Boolean functions: noise stability, non-interactive correlation, and mutual information

Jiange Li, Muriel Médard *

June 18, 2018

Abstract

Let $T_\epsilon$ be the noise operator acting on Boolean functions $f : \{0, 1\}^n \mapsto \{0, 1\}$, where $\epsilon \in [0, 1/2]$ is the noise parameter. Given $\alpha \geq 1$ and the mean $E f$, which Boolean function $f$ maximizes the $\alpha$-th moment $E(T_\epsilon f)^\alpha$? Our findings are: in the weak noise scenario, i.e., $\epsilon$ is small, the maximum is achieved by the lexicographic function; in the strong noise scenario, i.e., $\epsilon$ is close to 1/2, the maximum is achieved by Boolean functions with the largest degree-1 Fourier weight; and when $\alpha$ is a large integer, among balanced Boolean functions, the maximum is achieved by any function which is 0 on all strings with fewer than $n/2$ 1’s. Moreover, for any convex function $\Phi$, we show that the maximum of $E \Phi(T_\epsilon f)$ is achieved by some monotone function. Analogous results are established in more general contexts, such as Boolean functions defined on the discrete torus $(\mathbb{Z}/p\mathbb{Z})^n$, as well as noise stability in a tree model. We also discuss the relationships between this noise stability problem and the problem of non-interactive correlation distillation, as well as Courtade-Kumar’s conjecture on the most informative Boolean function.

1 Introduction

Let $T_\epsilon$ be the noise operator (the definition will be given in Section 2) acting on Boolean functions defined on the discrete cube $\{0, 1\}^n$ associated with the uniform measure. We are interested in the problem that given $\alpha \geq 1$ and the mean $E f$ which Boolean function $f$ maximizes the $\alpha$-th moment $E(T_\epsilon f)^\alpha$. The second moment, i.e., $\alpha = 2$, is related to the noise stability of $f$, and the study of noise stability plays an important role in many areas of mathematics and computer sciences, such as inapproximability [10], learning theory [3, 18], hardness amplification [24], mixing of short random walks [16], and percolation [4]. We may refer to $E(T_\epsilon f)^\alpha$ as $\alpha$-stability of the Boolean function $f$. In the Gaussian setting, Borell’s isoperimetric inequality [5] for two functions asserts that half spaces are the extremizers, and this is generalized by Isaksson and Mossel [14] to multiple functions. The general $\alpha$-stability in the Gaussian setting is studied by Eldan [7]. In the discrete setting, we still lack the knowledge of the extremizers.

The problem of noise stability is closely related to the problem of non-interactive correlation distillation (NICD) [21, 22, 29], which is relevant for cryptographic information reconciliation, random beacons in cryptography and security, and coding theory. In its most basic form, the problem of NICD involves two players. Let $X \in \{0, 1\}^n$ be a uniformly random binary string transmitted to Alice and Bob through independent binary symmetric channels.
(i.e., BSC(ε)) with cross-over probability ε, i.e., each bit of $X$ is flipped independently with probability ε. Suppose Alice and Bob receive $Y$ and $Y'$, respectively. They wish to maximize the agreement probability $\mathbb{P}(f(Y) = g(Y'))$ using Boolean functions $f$ and $g$, respectively. Notice that

$$
\mathbb{E}f(Y) = \mathbb{P}(f(Y) = g(Y') = 1) + \mathbb{P}(f(Y) = 1, g(Y') = 0),
$$

$$
\mathbb{E}g(Y') = \mathbb{P}(f(Y) = g(Y') = 1) + \mathbb{P}(f(Y) = 0, g(Y') = 1).
$$

We have

$$
\mathbb{P}(f(Y) = g(Y')) = 1 + 2\mathbb{E}f(Y)g(Y') - \mathbb{E}f(Y) - \mathbb{E}g(Y').
$$

Given $\mathbb{E}f$ and $\mathbb{E}g$, it suffices to maximize the correlation $\mathbb{E}f(Y)g(Y') = \mathbb{E}(T_\epsilon f T_\epsilon g)$. The goal of the $k$-player NICD problem is to maximize $\mathbb{P}(f_1(Y^1) = \cdots = f_k(Y^k))$, where $Y^1, \cdots, Y^k$ are $k$ noise corrupted versions of $X$, and $f_1, \cdots, f_k$ are Boolean functions. In general, there is no equivalence of the NICD problem and the problem of maximal correlation $\mathbb{E} \prod_{i=1}^k f_i(Y^i)$ = $\mathbb{E} \prod_{i=1}^k T_\epsilon f_i$. In certain scenarios, the maximizers of the correlation problem also play the extremal role in NICD problem. This will be discussed in more details in Section 3.

Let $X$ and $Y$ be random binary strings defined as before, and let $f$ be a Boolean function. Courtade and Kumar [6] conjectured that the mutual information $I(X; f(Y))$ between $X$ and $f(Y)$ is maximized by the dictator function. Its Gaussian analogy has recently been solved by Kindler, O’Donnell and Witmer [19]. Pichler, Piantanida and Matz [27] proved the variant that the dictator function maximizes the mutual information between $f(X)$ and $g(Y)$ among all Boolean functions $f$ and $g$. The original conjecture is only verified in the strong noise regime, i.e., $\epsilon$ close to 1/2, by Samorodnitsky [28]. Our observation is that Courtade-Kumar’s conjecture for balanced Boolean functions is equivalent to that the dictator function maximizes the $\alpha$-stability $\mathbb{E}(T_\epsilon f)^\alpha$ for $\alpha$ close to 1. This may provide a different perspective to study Courtade-Kumar’s conjecture.

The paper is organized as follows. In Section 2 we give a brief account of the noise operator and other notions used in the study of Boolean functions. We refer the interested reader to the monograph [25] for further information. In Section 3 we include results in different scenarios, such as the weak noise (i.e., $\epsilon$ is small) and the strong noise (i.e., $\epsilon$ is close to 1/2), as well as the asymptotic result when $\alpha$ is a large integer. In Section 4 we relate the problem of noise stability to Courtade-Kumar’s conjecture on the most informative Boolean function. In Section 5 we establish analogous results in general contexts, such as Boolean functions defined on the discrete torus $(\mathbb{Z}/p\mathbb{Z})^n$, as well as noise stability in a tree model. We conclude the paper with the discussion of potential applications and future work in Section 6.

2 Noise operator

Let $\{0,1\}^n$ be the discrete cube associated with the uniform measure $\mu$. It is known that $\{W_A(x) = (-1)^{\sum_{i \in A} x_i}\}_{A \subseteq [n]}$ forms an orthonormal basis, i.e., $\mathbb{E}(W_A)^2 = 1$ and $\mathbb{E}(W_A W_B) = 0$ for $A \neq B$. (The expectation is taken with respect to the reference measure $\mu$. We always omit it when it is clear from the context). Any real-valued function $f$ on $\{0,1\}^n$ has the following Fourier expansion

$$
f(x) = \sum_A \hat{f}(A) W_A(x),
$$

where $\hat{f}(A) = \mathbb{E}(f W_A)$ are the Fourier coefficients. Particularly, we have $\hat{f}(\emptyset) = \mathbb{E}f$.  

2
Definition 2.1. Let $0 \leq \epsilon \leq 1/2$. The noise operator $T_\epsilon$ acts on $f : \{0,1\}^n \mapsto \mathbb{R}$ as follows
\[ T_\epsilon f(x) = \mathbb{E} f(x + Z), \]
where $Z$ has independent Bernoulli($\epsilon$) coordinates, and the addition is mod by 2.

Let $X \in \{0,1\}^n$ be a uniform binary string, and let $Y$ be the output of $X$ through a BSC($\epsilon$) channel, i.e., $Y = X + Z$. Then we have $T_\epsilon f(x) = \mathbb{E}[f(Y)|X=x]$. Using the Fourier expansion, we have
\[ T_\epsilon f(x) = \sum_A (1 - 2\epsilon)|A|\hat{f}(A)W_A(x). \]

The noise operator $T_\epsilon$ has an smoothing effect; that is, the function $T_\epsilon f$ becomes “smoother” as $\epsilon$ grows. In particular, we have $T_0 f = f$ and $T_{1/2} f = \mathbb{E} f$.

Let $1 \leq i \leq k$. We define $Y^i = X + Z^i$, where $Z^i$ are independent copies of $Z$. Each pair $(Y^i, Y^j)$ for $i \neq j$ has the correlation matrix $\rho I$, where $\rho = (1 - 2\epsilon)^2$, and $I$ is the identity matrix. For simplicity, we say that they are $\rho$-correlated. Notice that $Y_i$ are independent given the information of $X$. Together with [1], the conditioning argument yields
\[ \mathbb{E} \prod_{i=1}^k f_i(Y^i) = \mathbb{E} \prod_{i=1}^k T_\epsilon f_i. \]

Owing to this relation, the results in the next section will be stated in terms of either LHS or RHS of the identity above.

The noise operator introduced before can be thought of as a special type of Markov semi-groups of Markov chains on graphs. (In our case, the graph is the discrete cube $\{0,1\}^n$.) To be more precise, let us consider the following simple continuous time Markov chain on a simple connected undirected graph $G = (V, E)$. Each vertex $x \in V$ is associated with an exponential clock, i.e., an exponential random variable with parameter 1. When the clock rings, the chain jumps to the neighbors of the current location with equal probability. In this case, the transition matrix is $K = D^{-1}A$, where $A$ is the adjacency matrix, and $D$ is the diagonal matrix with $D(x,x) = d_x$ the degree of $x$. The invariant measure of the Markov chain is $\mu(x) = d_x/\sum_y d_y$. The Markov semi-group $(P_t)_{t \geq 0}$ acts on $f : V \mapsto \mathbb{R}$ as follows
\[ P_t f(x) = e^{-tL} f(x), \]
where $L = I - K$ is the Laplacian. Differentiating the equation $\mathbb{E}(P_t f) = \mathbb{E} f$ with respect to $t$ at $t = 0$, we have $\mathbb{E}(Lf) = 0$ for any function $f$. By Jensen’s inequality, we have $P_t(\Phi(f)) \geq \Phi(P_t f)$ for convex functions $\Phi$. Differentiating this inequality with respect to $t$ at $t = 0$, we have $L(\Phi(f)) \geq \Phi'(f)Lf$. Therefore, we have
\[ \frac{d}{dt}\mathbb{E}\Phi(P_t f) = \mathbb{E}\Phi'(P_t f)L(P_t f) \leq \mathbb{E}L(\Phi(P_t f)) = 0, \]
i.e., $\mathbb{E}\Phi(P_t f)$ is a decreasing function of $t$.

An important notation used in the study of Boolean functions is the influence. The flipping operator $\sigma_i$ is defined as
\[ \sigma_i(x_1, \ldots, x_i, \ldots, x_n) = (x_1, \ldots, 1 - x_i, \ldots, x_n). \]
Definition 2.2. Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. The influence of the $i$-th variable $I_i(f)$ is defined as

$$I_i(f) = \mu(x : f(x) \neq f(\sigma_i(x))).$$

The total influence $I(f)$ is defined as

$$I(f) = \sum_{i=1}^{n} I_i(f).$$

We have the following geometrical interpretation of influence in terms of edge boundary. Let $S$ be the support of $f$. The $i$-th direction edge boundary $\partial_i S$ is defined as

$$\partial_i S = \{(x, \sigma_i(x)) : x \in S, \sigma_i(x) \notin S\}.$$  

Two vertices $x, y \in \{0,1\}^n$ are called adjacent, i.e., $x \sim y$, if and only if their Hamming distance is 1. The edge boundary $\partial S$ is defined as

$$\partial S = \{(x, y) : x \sim y, x \in S, y \notin S\}.$$  

It is easy to see that $\partial S = \cup_{i=1}^{n} \partial_i S$. One can check the following identities

$$I_i(f) = \frac{|\partial_i S|}{2^{n-1}}, \quad (3)$$

$$I(f) = \frac{|\partial S|}{2^{n-1}}. \quad (4)$$

We also have the following Fourier analytic representation of influence. Since $f$ takes values 0 or 1, one can rewrite $I_i(f)$ as

$$I_i(f) = \mathbb{E}(f(X) - f(\sigma_i(X)))^2,$$

where $X \in \{0,1\}^n$ is a uniform binary string. Using the Fourier expansion of $f$, we have

$$I_i(f) = \mathbb{E}\left(2 \sum_{A \ni i} \hat{f}(A) W_A(x)\right)^2 = 4 \sum_{A \ni i} \hat{f}(A)^2.$$  

Hence,

$$I(f) = 4 \sum_{A} |A| \hat{f}(A)^2.$$

The following are some typical Boolean functions needed in Section 3.

Definition 2.3. A lexicographic set is a set with elements selected in the lexicographic order. A Boolean function $f$ is called lexicographic if it is supported on a lexicographic set.

Definition 2.4. The majority function of $n$ (odd number) variables is defined as $\text{Maj}_n(x) = \text{sgn}(\sum_{i=1}^{n} x_i - n/2)$. The dictator function is $\text{Maj}_1$, which only looks at the first bit.

Definition 2.5. The natural partial order relation on $\{0,1\}^n$ is defined as $x \preceq y$ if $x_i \leq y_i$ holds for all $i \in [n]$. The function $f$ is called monotone if $f(x) \leq f(y)$, whenever $x \preceq y$, and $f$ is called anti-monotone if $f(x) \geq f(y)$, whenever $x \preceq y$. 

3 Main results

For the problem of $k$-player correlation, the following statement asserts that the players should use the same strategy to maximize their correlation.

**Proposition 3.1.** Let $0 < \rho < 1$. Let $Y^1, \cdots, Y^k \in \{0, 1\}^n$ be $\rho$-correlated uniform binary strings. For any functions $f_i : \{0, 1\}^n \mapsto \mathbb{R}$, we have

$$\mathbb{E} \prod_{i=1}^{k} f_i(Y^i) \leq \max_{1 \leq i \leq k} \mathbb{E} \prod_{j=1}^{n} f_i(Y^j).$$

Equality is achieved if and only if $f_i$ are multiples of the same function.

**Proof.** As we have seen, we can realize $Y^i$ as $X + Z^i$, where $X$ is a uniform binary sequence, and the coordinates of $Z^i$ are i.i.d. Bernoulli($\epsilon$) with $\epsilon = (1 - \sqrt{\rho})/2$. Since $Y^i$ are independent given $X$, we have

$$\mathbb{E} \prod_{i=1}^{k} f_i(Y^i) = \mathbb{E} \prod_{i=1}^{k} \mathbb{E}[f_i(Y^i)|X] \leq \prod_{i=1}^{k} (\mathbb{E}[f_i(Y^i)|X])^{1/k} \leq \max_{1 \leq i \leq k} \mathbb{E} \prod_{j=1}^{n} f_i(Y^j).$$

The first inequality follows from Hölder’s inequality, and equality is achieved if only if $\mathbb{E}[f_i(Y^i)|X]$ are multiples of the same function. Since the noise operator is invertible, $f_i$ should also be multiples of the same function.

A well known result of Harper [9] asserts that the lexicographical set has the least edge boundary among all subsets of $\{0, 1\}^n$ with a fixed size. Owing to the connection between the total influence and the size of the edge boundary, Harper’s theorem is equivalent to that the lexicographic function has the minimal total influence among all Boolean functions with the fixed mean.

**Theorem 3.1.** Let $\alpha > 1$, $n$ and $\mathbb{E} f$ be fixed. When $\epsilon = \epsilon(n)$ is sufficiently small, $\mathbb{E}(T_\epsilon f)^\alpha$ is maximized by the lexicographic function. When $\epsilon = \epsilon(n)$ is sufficiently close to 1/2, $\mathbb{E}(T_\epsilon f)^\alpha$ is maximized by Boolean functions with the maximal degree-1 Fourier weight $\sum_{i=1}^{n} \hat{f}(\{i\})^2$. If $f$ is assumed to be balanced, i.e., $\mathbb{P}(f = 0) = \mathbb{P}(f = 1)$, the dictator function maximizes $\mathbb{E}(T_\epsilon f)^\alpha$ in both scenarios.

**Proof.** Differentiating $\mathbb{E}(T_\epsilon f)^\alpha$ with respect to $\epsilon$, we have

$$\frac{d}{d\epsilon} \mathbb{E}(T_\epsilon f)^\alpha = 2\alpha(1 - 2\epsilon)^{-1} \mathbb{E}(T_\epsilon f)^{\alpha-1} (L \circ T_\epsilon) f,$$
where the operator $L$ is defined as
\[ Lf(x) = -\sum_A |A| \hat{f}(A)W_A(x). \quad (5) \]

In particular, we have
\[
\frac{d}{d\epsilon} \mathbb{E}(T_{\epsilon} f)^\alpha \bigg|_{\epsilon=0} = 2\alpha \mathbb{E}(f^{p-1} Lf) \\
= 2\alpha \mathbb{E}(f Lf) \\
= -2\alpha \sum_{A \subseteq [n]} |A| \hat{f}(A)^2 \\
= -\alpha I(f)/2, \quad (6)
\]

where $I(f)$ is the total influence of $f$ (given in Definition 2.2). Harper’s theorem [9] and (6) imply that $\frac{d}{d\epsilon} \mathbb{E}(T_{\epsilon} f)^\alpha \bigg|_{\epsilon=0}$ is maximized by the lexicographic function. For $\alpha > 1$, we know that $\mathbb{E}(T_{\epsilon} f)^\alpha$ is decreasing as a function of $\epsilon$. Since $\mathbb{E}(T_0 f)^\alpha = \mathbb{E} f$ is fixed, the lexicographic function maximizes $\mathbb{E}(T_{\epsilon} f)^\alpha$ for small $\epsilon > 0$. When $\epsilon$ is close to $1/2$,
\[
T_{\epsilon} f(x) = \mathbb{E} f + (1 - 2\epsilon) \sum_{i=1}^n \hat{f}(\{i\})(-1)^x_i + O((1 - 2\epsilon)^2),
\]
where $O(\cdot)$ depends on $n, k$ not $f$. Then we have
\[
(T_{\epsilon} f(x))^{\alpha-1} = (\alpha - 1)(\mathbb{E} f)^{\alpha-2}(1 - 2\epsilon) \sum_{i=1}^n \hat{f}(\{i\})(-1)^x_i + (\mathbb{E} f)^{\alpha-1} + O((1 - 2\epsilon)^2),
\]
and
\[
(L \circ T_{\epsilon}) f(x) = -(1 - 2\epsilon) \sum_{i=1}^n \hat{f}(\{i\})(-1)^x_i + O((1 - 2\epsilon)^2).
\]

Since $O(\cdot)$ term has mean zero, for $\epsilon$ close to $1/2$, we have
\[
\frac{d}{d\epsilon} \mathbb{E}(T_{\epsilon} f)^\alpha = -2\alpha(\alpha - 1)(1 - 2\epsilon)(\mathbb{E} f)^{\alpha-2} \sum_{i=1}^n \hat{f}(\{i\})^2 + O((1 - 2\epsilon)^2).
\]

Notice that $\mathbb{E}(T_{1/2} f)^p = (\mathbb{E} f)^\alpha$. When $\epsilon$ is sufficiently close to $1/2$, $\mathbb{E}(T_{\epsilon} f)^\alpha$ is maximized by the function which maximizes $\sum_{i=1}^n \hat{f}(\{i\})^2$. If $f$ is assumed to be balanced, it is clear that lexicographic function is just the dictator function. Notice that
\[
\sum_{i=1}^n \hat{f}(\{i\})^2 \leq \sum_{A \subseteq [n]} \hat{f}(A)^2 = \mathbb{E} f - (\mathbb{E} f)^2
\]
and that equality holds if $f$ is the dictator function. \qed

**Remark 3.2.** Our arguments could yield explicit bounds of the noise level $\epsilon$, but they will depend on the dimension $n$. It is reasonable to ask if one can make it dimension-free.
If $\mathbb{E}(T_{\epsilon} f)^2$ is maximized by the lexicographic function, it also achieves the maximum at the function supported on a set in the reverse lexicographic order. This implies that $\mathbb{E}(T_{\epsilon} f)^{\alpha}$ and $\mathbb{E}(T_{\epsilon} (1 - f))^\alpha$ are maximized by the same lexicographic function. It is clear that $f$ and $1 - f$ have the same degree-1 Fourier weight. As a consequence of Theorem 3.1 we have the following result on the $k$-player NICD problem, which is proved by Mossel and O’Donnell [21] for balanced Boolean functions.

**Corollary 3.3.** Let $0 < \rho < 1$. Let $Y_1, \cdots, Y_k \in \{0,1\}^n$ be $\rho$-correlated uniform binary strings. If $\rho$ is sufficiently small, then the agreement probability $\mathbb{P}(f(Y_1) = \cdots = f(Y_k))$ is maximized by the Boolean function $f$ with the maximal degree-1 Fourier weight $\sum_{i=1}^n \hat{f}(\{i\})^2$.

In the strong correlation regime, i.e., $\rho$ close to 1, we have the following heuristic explanation for the two-player case. Suppose that $f$ is supported on $S$. Our goal is to maximize $\mathbb{P}(Y_1 \in S, Y_2 \in S)$, which is equivalent to minimize $\mathbb{P}(Y_1 \in S, Y_2 \notin S)$. Since $Y_1$ and $Y_2$ are $\rho$-correlated, we can think of $Y_2$ as obtained from $Y_1$ by flipping its coordinates independently with probability $(1 - \rho)/2$. When $\rho$ is close to 1, it is more likely that $Y_1$ and $Y_2$ only differ by one bit, i.e., $(Y_1, Y_2)$ belongs to the edge boundary. Then smaller edge boundary implies larger agreement probability. Harper’s theorem [9] asserts that the lexicographic set has the least boundary among all sets with fixed size.

The question of precisely maximizing the degree-1 Fourier weight among Boolean functions with fixed mean is a well-known difficult one. The following example shows that the indicator of a Hamming ball is superior to the lexicographic function when the mean is sufficiently small. Suppose that $E f = 2^{-m}$ for $1 \leq m \leq n$. The lexicographic function $f(x) = \prod_{i=1}^m x_i$ is supported on a sub-cube $S$. Let $\rho = (1 - 2\epsilon)^2$. We have

$$\mathbb{P}(f(Y_1) = f(Y_2)) = 4^{-n}(1 + \rho)^m |S|.$$  

When $|S|$ is small, we can let $g$ be a Boolean function supported on a vertex and $|S| - 1$ vertices with Hamming distance 1 from that vertex. Elementary calculations yield

$$\mathbb{P}(g(Y_1) = g(Y_2)) = 4^{-n}(1 + \rho)^{n-2}[(1 - \rho)^2 |S|^2 + 4\rho(2 - \rho)|S| - 4\rho].$$

Then we have

$$\frac{d}{d\rho} \mathbb{P}(f(Y_1) = f(Y_2)) \big|_{\rho=0} = 4^{-n}m|S|^2,$$

$$\frac{d}{d\rho} \mathbb{P}(g(Y_1) = g(Y_2)) \big|_{\rho=0} = 4^{-n}[(n - 4)|S|^2 + 8|S| - 4].$$

For $n - \log_2 n < m \leq n - 4$, we have

$$\frac{d}{d\rho} \mathbb{P}(f(Y_1) = f(Y_2)) \big|_{\rho=0} < \frac{d}{d\rho} \mathbb{P}(g(Y_1) = g(Y_2)) \big|_{\rho=0}.$$

This implies that $\mathbb{P}(f(Y_1) = f(Y_2)) < \mathbb{P}(g(Y_1) = g(Y_2))$ for small $\rho > 0$.

Among balanced functions, the dictator function maximizes $\mathbb{E}(T_{\epsilon} f)^2$ at any noise level. This simply follows from

$$\mathbb{E}(T_{\epsilon} f)^2 = \sum_{A \subseteq [n]} (1 - 2\epsilon)^{2|A|} \hat{f}(A)^2 \leq (\mathbb{E} f)^2 + (1 - 2\epsilon)^2(\mathbb{E} f - (\mathbb{E} f)^2).$$
Equality is achieved for the dictator function. Notice that for balanced functions
\[ \mathbb{E}f(Y^1)f(Y^2) = \mathbb{E}(1 - f(Y^1))(1 - f(Y^2)). \]
Therefore, we have
\[ \mathbb{P}(f(Y^1) = f(Y^2)) = 2\mathbb{E}f(Y^1)f(Y^2), \]
which is maximized by the dictator function. Similarly, we have
\[ \mathbb{P}(f(Y^1) = f(Y^2) = f(Y^3)) = 3\mathbb{E}f(Y^1)f(Y^2) - 1/2. \]
Therefore, the dictator function is still the best strategy in the three-player case. This recovers Theorem 1.3 of Mossel and O’Donnell \[21\]. However, we do not know if the dictator function also maximizes the third moment \( \mathbb{E}(T_{\epsilon}f)^3 \) among balanced Boolean functions.

**Theorem 3.2.** Let \( \Phi \) be a convex function. For fixed mean \( \mathbb{E}f \), the quantity \( \mathbb{E}\Phi(T_{\epsilon}f) \) is maximized by some monotone function.

**Proof.** The proof is inspired by a shifting technique in \[17\] and a convex combination argument in \[6\] (Theorem 3). Suppose that \( f \) is supported on \( S \). Let \( S'_2 \) be the projection of \( S \) on the last \( n - 1 \) bits, i.e., \( x^0_n \in S'_2 \) if \( (0, x^0_2) \in S \) or \( (1, x^0_2) \in S \). We define the following partition of \( S'_2 \):
\[
A = \{ x^0_n \in S'_2 : (0, x^0_2) \in S, (1, x^0_2) \in S \}, \\
B = \{ x^0_n \in S'_2 : (0, x^0_2) \in S, (1, x^0_2) \notin S \}, \\
C = \{ x^0_n \in S'_2 : (0, x^0_2) \notin S, (1, x^0_2) \in S \}.
\]
Then we have \( S = (\{0,1\} \times A) \cup (\{0\} \times B) \cup (\{1\} \times C) \). Let \( g \) be the Boolean function supported on \( S' = (\{0,1\} \times A) \cup (\{1\} \times \{B, C\}) \). It is clear that \( |S| = |S'| \), and that \( f \) and \( g \) have the same mean. We claim that \( g \) is superior to \( f \), i.e., \( \mathbb{E}_f(T_{\epsilon}f) \leq \mathbb{E}_f(T_{\epsilon}g) \). Let \( h \) be the Boolean function supported on \( S'' = (\{0,1\} \times A) \cup (\{0\} \times \{B, C\}) \). For any \( x \in \{0,1\}^n \), we will show that
\[
T_{\epsilon}f(x) = \theta T_{\epsilon}g(x) + (1 - \theta)T_{\epsilon}h(x),
\]
where \( \theta \) depends on \( x^0_n \). We only check this identity for \( x = (0, x^0_2) \), since the argument is similar for \( x = (1, x^0_2) \). Let \( X \in \{0,1\}^n \) be a uniformly random binary string. Let \( Y = X + Z \), where the coordinates of \( Z \) are i.i.d. Bernoulli(\( \epsilon \)). Then we have
\[
T_{\epsilon}f(0, x^0_2) = \mathbb{P}(f(Y) = 1|X = (0, x^0_2)) \\
= \mathbb{P}(Y_2^n \in A|X = (0, x^0_2)) + \mathbb{P}(Y_1 = 0, Y_2^n \in B|X = (0, x^0_2)) + \mathbb{P}(Y_1 = 1, Y_2^n \in C|X = (0, x^0_2)) \\
= \mathbb{P}(Y_2^n \in A|X = (0, x^0_2)) + \mathbb{P}(Y_1 = 1, Y_2^n \in B|X = (0, x^0_2)) + \mathbb{P}(Y_1 = 1, Y_2^n \in C|X = (0, x^0_2)) + (1 - 2\epsilon)\mathbb{P}(Y_2^n \in B|X_2^n = x^0_2) \\
= T_{\epsilon}g(0, x^0_2) + (1 - 2\epsilon)\mathbb{P}(Y_2^n \in B|X_2^n = x^0_2).
Similarly, we have
\[ T_\epsilon f(0, x^n_2) = T_\epsilon h(0, x^n_2) - (1 - 2\epsilon) \mathbb{P}(Y^n_2 \in C|X^n_2 = x^n_2). \]
Therefore, identity (7) holds with
\[ \theta = \frac{\mathbb{P}(Y^n_2 \in C|X^n_2 = x^n_2)}{\mathbb{P}(Y^n_2 \in B|X^n_2 = x^n_2) + \mathbb{P}(Y^n_2 \in C|X^n_2 = x^n_2)}. \]
Notice that \( \theta \) is independent of \( x_1 \). We first apply the convex function \( \Phi \) to (7), and then average both sides over the first bit. Then we have
\[
\mathbb{E}\Phi(T_\epsilon f(X_1, x^n_2)) \leq \theta \mathbb{E}\Phi(T_\epsilon g(X_1, x^n_2)) + (1 - \theta) \mathbb{E}\Phi(T_\epsilon h(X_1, x^n_2)),
\]
which follows from
\[
\begin{align*}
T_\epsilon g(0, x^n_2) &= \mathbb{P}(Y^n_2 \in A|X_1 = 0, X^n_2 = x^n_2) \\
&= \mathbb{P}(Y^n_2 \in A|X^n_2 = x^n_2) \\
&= T_\epsilon h(0, x^n_2).
\end{align*}
\]
Hence, inequality (8) becomes
\[
\mathbb{E}\Phi(T_\epsilon f(X_1, x^n_2)) \leq \mathbb{E}\Phi(T_\epsilon g(X_1, x^n_2)).
\]
We will have \( \mathbb{E}\Phi(T_\epsilon f) \leq \mathbb{E}\Phi(T_\epsilon g) \) by averaging both sides of the above inequality over \( x^n_2 \). Repeat the argument over the last \( n - 1 \) bits. We will arrive at a monotone function.

**Remark 3.4.** Theorem 3.2 was proved in [21] for \( \Phi(x) = x^k \) where \( k \) is a positive integer. In this case, the theorem can be rephrased in the following way. Let \( 0 < \rho < 1 \). Let \( Y^1, \ldots, Y^k \in \{0, 1\}^n \) be \( \rho \)-correlated uniform binary strings. Given the mean \( \mathbb{E}f \), the quantity \( \mathbb{E}\prod_{i=1}^k f(Y^i) \) is maximized by some monotone function.

If \( \mathbb{E}\prod_{i=1}^k f(Y^i) \) is maximized by some monotone function, it should also achieve the maximum at an anti-monotone function. Therefore, \( \mathbb{E}\prod_{i=1}^k f(Y^i) \) and \( \mathbb{E}\prod_{i=1}^k (1 - f)(Y^i) \) can be maximized by the same monotone function. As a consequence of Theorem 3.2, we have the following result on the \( k \)-player NICD problem, which was obtained by Mossel and O’Donnell [21] for balanced Boolean function.

**Corollary 3.5.** Let \( 0 < \rho < 1 \). Let \( Y^1, \ldots, Y^k \in \{0, 1\}^n \) be \( \rho \)-correlated uniform binary strings. Given the mean \( \mathbb{E}f \), the agreement probability \( \mathbb{P}(f(Y^1) = \cdots = f(Y^k)) \) is maximized by some monotone function.
We have seen from Theorem 3.1 that among balanced Boolean functions the dictator function maximizes $E(T_\epsilon f)^k$ in both the weak noise and the strong noise scenarios for fixed $n$ and $k$. One may expect that the same property holds for arbitrary noise. The following result asserts that this is not true if $k$ is large.

**Theorem 3.3.** Let $n, \epsilon$ be fixed and let $k$ be sufficiently large. Among balanced Boolean functions, the quantity $E(T_\epsilon f)^k$ is maximized by any function which is 0 on all strings with fewer than $n/2$ 1’s. In particular, for $n$ odd, $E(T_\epsilon f)^k$ is maximized by the majority function $\text{Maj}_n$.

The proof depends on the following observations due to Mossel and O’Donnell [21].

**Lemma 3.6.** For monotone functions, $T_\epsilon f(x)$ is maximized at $x = \overline{1} = (1, \cdots, 1)$.

**Proof.** Notice that flipping each bit of a string with probability $\epsilon$ is equivalent to update each bit with probability $2\epsilon$, where the update consists of replacing the bit by a random choice from $\{0, 1\}$. For any $x \in \{0, 1\}^n$, we denote by $x'$ the noised version of $x$. We couple the random sequences $x'$ and $\overline{1}$ in the following way: update the same bit of $x$ and $\overline{1}$ with the same value. It is clear that $x' \leq \overline{1}$. By monotonicity, we have $f(\overline{1}) = 1$ if $f(x') = 1$. Therefore, we have $T_\epsilon f(x) \leq T_\epsilon f(\overline{1})$. \hfill $\Box$

**Lemma 3.7.** Among balanced Boolean functions, $(T_\epsilon f)(\overline{1})$ is maximized by any function which is 0 on all strings with fewer than $n/2$ 1’s. In particular, for $n$ odd, $(T_\epsilon f)(\overline{1})$ is maximized by the majority function $\text{Maj}_n$.

**Proof.** The statement simply follows from

$$T_\epsilon f(\overline{1}) = \sum_{x \in S} \epsilon^{d(x, \overline{1})}(1 - \epsilon)^{n - d(x, \overline{1})},$$

where $S$ is the support of $f$, and $d(x, \overline{1})$ is the Hamming distance between $x$ and $\overline{1}$, and the simple fact that the quantity being summed is strictly decreasing with respect to $d(x, \overline{1})$. \hfill $\Box$

**Proof.** (Theorem 3.3.) We only prove the theorem for $n$ odd, since the proof for $n$ even is essentially the same. Invoke Theorem 3.6 then we can assume that $f$ is monotone. Using Lemma 3.6 we have

$$E(T_\epsilon f)^k = 2^{-n} \sum_x (T_\epsilon f(x))^k \leq (T_\epsilon f(\overline{1}))^k.$$

It is clear that

$$E(T_\epsilon \text{Maj}_n)^k = 2^{-n} \sum_x (T_\epsilon \text{Maj}_n(x))^k \geq 2^{-n}(T_\epsilon \text{Maj}_n(\overline{1}))^k.$$

By Lemma 3.7 if $f \neq \text{Maj}_n$, we have $T_\epsilon f(\overline{1}) < T_\epsilon \text{Maj}_n(\overline{1})$. For sufficiently large $k$, we will have $(T_\epsilon f(\overline{1}))^k < 2^{-n}(T_\epsilon \text{Maj}_n(\overline{1}))^k$. \hfill $\Box$

Then we can recover the following result of Mossel and O’Donnell [21].

**Corollary 3.8.** Let $0 < \rho < 1$. Let $Y^1, \cdots, Y^k \in \{0, 1\}^n$ be $\rho$-correlated uniform binary strings. For sufficiently large $k$, among balanced Boolean functions, $P(f(Y^1) = \cdots = f(Y^k))$ maximized by any function which is 0 on all strings with fewer than $n/2$ 1’s. For $n$ odd, the agreement probability is maximized by the majority function $\text{Maj}_n$. 

10
Hence, for fixed $\epsilon$, $n$ even and $1 < r < n$ such that the function $\text{Maj}_r(x) = \text{sgn} \left( \sum_{i=1}^{n} x_i - r/2 \right)$ is superior to both the dictator function and the majority function. Consider the numerical example $k = 10, \epsilon = 0.26, n = 5, r = 3$, which is taken from \[21\] (Proposition 5.2). One can check that $\mathbb{E}(T, \text{Maj}_1)^{10} \leq 0.0247, \mathbb{E}(T, \text{Maj}_5)^{10} \leq 0.0244$ and $\mathbb{E}(T, \text{Maj}_3)^{10} \geq 0.0248$. We do not know whether $\mathbb{E}(T, f)^k$ is always maximized by some $\text{Maj}_r$.

4 The most informative Boolean function

Let $X \in \{0, 1\}^n$ be a uniform binary string, and let $Y$ be the output of $X$ through a BSC($\epsilon$) channel, i.e., $Y = X + Z$, where the coordinates of $Z$ are i.i.d. Bernoulli($\epsilon$). Let $f$ be a Boolean function. Courtade and Kumar \[6\] conjectured that the dictator function maximizes the mutual information $I(X; f(Y))$ between $X$ and $f(Y)$. This is known as the most informative Boolean function problem.

Owing to the fact that $f$ is Boolean, we have

$$I(X; f(Y)) = H(\mathbb{E}f) - H(f(Y)|X).$$

Hence, for fixed $\mathbb{E}f$, it suffices to maximize

$$-H(f(Y)|X) = \mathbb{E}[T_{\epsilon} f \log T_{\epsilon} f + T_{\epsilon} (1 - f) \log T_{\epsilon} (1 - f)].$$

This follows from $T_{\epsilon} f(X) = \mathbb{E}[f(Y)|X]$ and the Booleanity of $f$. Using \[9\], we have

$$-H(f(Y)|X) = \frac{d}{d\alpha} \mathbb{E}[(T_{\epsilon} f)^\alpha + (T_{\epsilon} (1 - f))^\alpha] \bigg|_{\alpha = 1}.$$

The initial value $\mathbb{E}[T_{\epsilon} f + T_{\epsilon} (1 - f)] = 1$ is fixed. Hence, for fixed $\mathbb{E}f$ and $\alpha$ close to 1, the maximizer of $\mathbb{E}[(T_{\epsilon} f)^\alpha + (T_{\epsilon} (1 - f))^\alpha]$ also maximizes $I(X; f(Y))$, and vice versa. Recall our discussion of the NICD problem (the paragraph before Corollary 3.3). As a consequence of Theorem 3.1, we have

**Corollary 4.1.** Let $\mathbb{E}f$ be fixed. When $\epsilon = \epsilon(n)$ is sufficiently small, the mutual information $I(X; f(Y))$ is maximized by the lexicographic function. When $\epsilon = \epsilon(n)$ is sufficiently close to 1/2, the mutual information $I(X; f(Y))$ is maximized by Boolean functions with the largest degree-1 Fourier weight $\sum_{i=1}^{n} \hat{f}(\{i\})^2$. In particular, within the class of balanced Boolean functions, the dictator function maximizes the mutual information $I(X; f(Y))$ in both scenarios.

**Remark 4.2.** Courtade-Kumar’s conjecture is verified by Ordentlich, Shayeavitz, and Weinstein \[26\] in the extremal settings: $\epsilon = \epsilon(n) \approx 0$ and $\epsilon = \epsilon(n) \approx 1/2$. Samorodnitsky \[28\] gave a dimension-free bound on the noise level in the strong noise setting. Our result provides a finer characterization of the maximizer of the mutual information $I(X; f(Y))$ under the constraint that the mean $\mathbb{E}f$ is fixed.

**Remark 4.3.** As a consequence of Theorem 3.1, it suffices to study Courtade-Kumar’s conjecture for monotone functions. This was also observed by Huleihel and Ordentlich \[11\].
When $\alpha = 1$ and $\alpha = 2$, the dictator function is the maximizer of $\mathbb{E}[(T_\alpha f)^\alpha + (T_\alpha(1 - f))^\alpha]$ within the class of balanced Boolean functions. It is reasonable to expect that the dictator function still plays the extremal role for any $1 < \alpha < 2$. So we conjecture that the following statement holds, which in particular implies Courtade-Kumar’s conjecture for balanced Boolean functions.

**Conjecture 4.4.** For $1 \leq \alpha \leq 2$, the dictator function maximizes $\mathbb{E}(T_\alpha f)^\alpha$ within the class of balanced Boolean functions.

**Remark 4.5.** Let $\Phi$ be a convex function. The $\Phi$-entropy of a function $f : \{0, 1\}^n \mapsto \mathbb{R}$ is defined as $H_\Phi(f) = \mathbb{E}\Phi(f) - \Phi(\mathbb{E}f)$. Let $\Phi(x) = 1 + x \log x + (1 - x) \log(1 - x)$. Then Courtade-Kumar’s conjecture can be rephrased as that the dictator function maximizes $H_\Phi(T_\alpha f)$ among all Boolean functions. We considered the function $\Phi(x) = x^\alpha$ for $1 < \alpha < 2$, and conjectured that the dictator function is the maximizer of $\Phi$-entropy $H_\Phi(T_\alpha f)$ within the class of balanced Boolean functions. Anantharam et al. [1] conjectured that the dictator function is still the maximizer for $H_\Phi(T_\alpha f)$ with the convex function $\Phi(x) = 1 - 2\sqrt{x(1 - x)}$, which is the squared Hellinger distance between two Bernoullis with parameters $x$ and $1 - x$, respectively.

## 5 General models

In this section, we discuss extensions of the problem of noise stability in two distinct ways: one is that the message transmitted is a multi-alphabet sequence rather than a binary sequence; the other is the problem of noise stability in a network in terms of a tree, which gives the geometry of the problem.

### 5.1 Discrete torus

In this model, we study the noise stability of Boolean functions defined on the discrete torus $(\mathbb{Z}/p\mathbb{Z})^n$, where $\mathbb{Z}/p\mathbb{Z} = \{0, 1, \cdots, p-1\}$ is the cyclic group of order $p$. As in the discrete cube case, our analysis of noise stability will rely on the combinatorial/graphic or group structure of the discrete torus.

Firstly, we give a brief introduction of Fourier analysis on the discrete torus $(\mathbb{Z}/p\mathbb{Z})^n$ associated with the uniform measure $\mu$. We denote by $e_p(\theta) = e^{2\pi i \theta/p}$. One can check that $\{e_p(\xi \cdot x)\}_{\xi \in (\mathbb{Z}/p\mathbb{Z})^n}$ forms an orthonormal basis, where $\xi \cdot x = \xi_1 x_1 + \cdots + \xi_n x_n$. Any function $f : (\mathbb{Z}/p\mathbb{Z})^n \mapsto \mathbb{R}$ has the following Fourier representation

$$f(x) = \sum_{\xi} \hat{f}(\xi)e_p(\xi \cdot x),$$

where the Fourier coefficients $\hat{f}(\xi) = \mathbb{E}f(x)e_p(\xi \cdot x)$.

Here is one way to define the noise operator on the discrete torus. Let $X = (X_1, \cdots, X_n) \in (\mathbb{Z}/p\mathbb{Z})^n$ be a uniform random vector; that is, the coordinates $X_j$ are independent and uniform on $\mathbb{Z}/p\mathbb{Z}$. The noise vector $Z = (Z_1, \cdots, Z_n) \in (\mathbb{Z}/p\mathbb{Z})^n$ consists of independent coordinates $Z_j$ with the same distribution, which assigns probability $1 - \epsilon$ to 0 and probability $\epsilon/(p-1)$ to any other elements. Let $Y = X + Z$.

**Definition 5.1.** Let $0 \leq \epsilon \leq 1 - 1/p$. The noise operator $T_\epsilon$ acting on $f : (\mathbb{Z}/p\mathbb{Z})^n \mapsto \mathbb{R}$ is defined as

$$T_\epsilon f(x) = \mathbb{E}f(x + Z).$$
Employing the identity $\sum_{j=0}^{p-1} e_p(jk) = 0$ for any $k \neq 0$, one can check that

$$T_\epsilon f(x) = \sum_\xi \left( 1 - \frac{p\epsilon}{p-1} \right)^{|\text{supp}(\xi)|} \hat{f}(\xi) e_p(\xi \cdot x),$$

(10)

where supp$(\xi) = \{ j : \xi_j \neq 0 \}$. When $p = 2$, this coincides with Definition 2.1.

Correspondingly, the NICD problem can be stated in the following way. We select a random sequence $x = (x_1, \cdots, x_n) \in (\mathbb{Z}/p\mathbb{Z})^n$, and pass it on to $k$ players through independent memoryless noise channels, which preserve the value of each $x_j$ with probability $1 - \epsilon$ and changes the value of $x_j$ to other values equally likely. Upon receiving the message, each player applies a Boolean function to output one alphabet. As usual, their goal is to maximize the agreement probability. We let $Y_1, \cdots, Y_k$ denote $k$ corrupted versions of $X$. The problem of NICD is to maximize the agreement probability $\mathbb{P}(f_1(Y_1) = \cdots = f_k(Y_k))$ using Boolean functions $f_1, \cdots, f_k$.

One can check that Proposition 3.1 still holds in this multi-alphabet setting, i.e., the $k$ players should apply the same Boolean function. Then the problem of NICD really asks for which Boolean function maximizes the $\alpha$-stability $E(T_\epsilon f)^\alpha$ with $\alpha = k$.

We will show an analog of Theorem 3.1. We need redefine the notation of influence of a Boolean function.

Let $\tilde{Z}_j$ be a random variable uniform on $\mathbb{Z}_p \setminus \{0\}$. The random flipping operator $\tilde{\sigma}_j$ is defined as

$$\tilde{\sigma}_j(x) = (x_1, \cdots, x_j + \tilde{Z}_j, \cdots, x_n).$$

Definition 5.2. Let $f : (\mathbb{Z}/p\mathbb{Z})^n \mapsto \{0, 1\}$ be a Boolean function. The influence of the $j$-th variable $I_j(f)$ is defined as

$$I_j(f) = \mathbb{P}(f(X) \neq f(\tilde{\sigma}_j(X))).$$

(We assume that $\tilde{Z}_j$ is independent of $X$). The total influence $I(f)$ is defined as

$$I(f) = \sum_{j=1}^n I_j(f).$$

Proposition 5.3. Let $f : (\mathbb{Z}/p\mathbb{Z})^n \mapsto \{0, 1\}$ be a Boolean function. Then we have

$$I_j(f) = \frac{2p}{p-1} \sum_{\xi : \xi_j \neq 0} |\hat{f}(\xi)|^2,$$

and

$$I(f) = \frac{2p}{p-1} \sum_\xi |\text{supp}(\xi)| |\hat{f}(\xi)|^2.$$

Proof. Since $f$ takes values 0 or 1, one can rewrite $I_i(f)$ as

$$I_j(f) = \mathbb{E}(f(X) - f(\tilde{\sigma}_j(X))^2).$$

(11)

Notice that both $X$ and $\tilde{\sigma}_j(X)$ are uniform. By Parseval’s identity, we have

$$\mathbb{E}f(X)^2 = \mathbb{E}f(\tilde{\sigma}_j(X))^2 = \sum_{\xi} |\hat{f}(\xi)|^2.$$

(12)
Using the Fourier representation, we have

$$\mathbb{E} f(X)f(\tilde{\sigma}_j(X)) = \mathbb{E} \sum_{\xi, \eta} \hat{f}(\xi)\overline{\hat{f}(\eta)} e_p((\xi - \eta) \cdot X)e_p(-\eta_j \tilde{Z}_j),$$

where $\overline{\hat{f}(\eta)}$ is the conjugate of $\hat{f}(\eta)$. Since $\tilde{Z}_j$ and $X$ are independent, we have

$$\mathbb{E}e_p((\xi - \eta) \cdot X)e_p(-\eta_j \tilde{Z}_j) = \mathbb{E} e_p((\xi - \eta) \cdot X)\mathbb{E} e_p(-\eta_j \tilde{Z}_j).$$

Owing to the orthogonality, $\mathbb{E} e_p((\xi - \eta) \cdot X)$ vanishes if $\xi \neq \eta$. One can check that

$$\mathbb{E} e_p(-\eta_j \tilde{Z}_j) = \begin{cases} 1, & \eta_j = 0 \\ -\frac{1}{p-1}, & \eta_j \neq 0. \end{cases}$$

Therefore, we have

$$\mathbb{E} f(X)f(\tilde{\sigma}_j(X)) = \sum_{\xi: \xi_j = 0} |\hat{f}(\xi)|^2 - \frac{1}{p-1} \sum_{\xi: \xi_j \neq 0} |\hat{f}(\xi)|^2. \tag{13}$$

The desired statement follows from (11), (12) and (13). \hfill \Box

**Theorem 5.1.** Let $\alpha > 1$, $n$ and $\mathbb{E} f$ be fixed. When $\epsilon$ is sufficiently small, $\mathbb{E}(T_\epsilon f)^\alpha$ is maximized by a Boolean function with the least total influence. When $\epsilon$ is sufficiently close to $1 - 1/p$, $\mathbb{E}(T_\epsilon f)^\alpha$ is maximized by a Boolean function with the maximal degree-1 Fourier weight $\sum_{|\text{supp}(\xi)| = 1} |\hat{f}(\xi)|^2$.

**Proof.** The statement can be proved in a manner similar to that of Theorem 3.1. We only give a sketch. In the low noise case, the equation

$$\frac{d}{d\epsilon} \mathbb{E}(T_\epsilon f)^\alpha |_{\epsilon = 0} = -\alpha I(f)/2$$

still holds with the total influence $I(f)$ given the Definition 5.2. When $\epsilon$ is close to $1 - 1/p$, one can check that the leading term of $\frac{d}{d\epsilon} \mathbb{E}(T_\epsilon f)^\alpha$ is

$$-\alpha(\alpha - 1)(\mathbb{E} f)^{\alpha-2} \frac{p}{p-1} \left( 1 - \frac{p\epsilon}{p-1} \right) \sum_{\xi: |\text{supp}(\xi)| = 1} |\hat{f}(\xi)|^2.$$

\hfill \Box

We have the following analogy of Theorem 5.2.

**Theorem 5.2.** Let $\Phi$ be a convex function. For fixed mean $\mathbb{E} f$, the quantity $\mathbb{E}\Phi(T_\epsilon f)$ is maximized by some monotone function.

**Proof.** We only need to slightly modify the proof of Theorem 3.2. Suppose that $f$ is supported on $S$. For each pair $j, k \in \mathbb{Z}/p\mathbb{Z}$ such that $j < k$, we define

$$B_{j,k} = \{ x^n_2 \in (\mathbb{Z}/p\mathbb{Z})^{n-1} : (j, x^n_2) \in S, (k, x^n_2) \notin S \},$$

and

$$C_{j,k} = \{ x^n_2 \in (\mathbb{Z}/p\mathbb{Z})^{n-1} : (j, x^n_2) \notin S, (k, x^n_2) \in S \}.$$
Let $A_{j,k} = S \setminus (\{j\} \times B_{j,k}) \cup (\{k\} \times C_{j,k})$. Let $g_{j,k}$ be the Boolean function supported on $S'_{j,k} = A_{j,k} \cup (\{k\} \times (B_{j,k} \cup C_{j,k}))$. It is clear that $|S| = |S'_{j,k}|$, and that $f$ and $g_{j,k}$ have the same mean. We claim that $g_{j,k}$ is superior to $f$, i.e., $\mathbb{E}\varphi(Te.f) \leq \mathbb{E}\varphi(Te.g_{j,k})$. Let $h_{j,k}$ be the Boolean function with support $S''_{j,k} = A_{j,k} \cup (\{j\} \times (B_{j,k} \cup C_{j,k}))$. For any $x \in (\mathbb{Z}/p\mathbb{Z})^n$, the following identity still holds

$$T_e f(x) = \theta T_e g(x) + (1 - \theta) T_e h(x), \quad (14)$$

where $\theta$ depends on $x^n_2$. For $x = (j, x^n_2)$, we have

$$T_e f(x) = \mathbb{P}(Y \in A_{j,k} | X = (j, x^n_2)) + \mathbb{P}(Y_1 = j, Y^n_2 \in B_{j,k} | X = (j, x^n_2)) + \mathbb{P}(Y_1 = k, Y^n_2 \in C_{j,k} | X = (j, x^n_2))$$

$$= \mathbb{P}(Y \in A_{j,k} | X = (j, x^n_2)) + \mathbb{P}(Y_1 = k, Y^n_2 \in B_{j,k} | X = (j, x^n_2)) + (1 - \frac{p\epsilon}{p-1}) \mathbb{P}(Y^n_2 \in B_{j,k} | X^n_2 = x^n_2)$$

$$= T_e g_{j,k}(x) + \left(1 - \frac{p\epsilon}{p-1}\right) \mathbb{P}(Y^n_2 \in B_{j,k} | X^n_2 = x^n_2).$$

Similarly, we have

$$T_e f(x) = T_e h_{j,k}(x) - \left(1 - \frac{p\epsilon}{p-1}\right) \mathbb{P}(Y^n_2 \in C_{j,k} | X^n_2 = x^n_2).$$

Therefore, identity (14) holds with

$$\theta = \frac{\mathbb{P}(Y^n_2 \in C_{j,k} | X^n_2 = x^n_2)}{\mathbb{P}(Y^n_2 \in B_{j,k} | X^n_2 = x^n_2) + \mathbb{P}(Y^n_2 \in C_{j,k} | X^n_2 = x^n_2)}.$$

The case $x = (k, x^n_2)$ can be checked in the same manner. When $x_1 \neq j, k$, we have $T_e f(x) = T_e g_{j,k}(x) = T_e h_{j,k}(x)$. Hence, we first apply the convex function $\Phi$ to (14), and then average both sides over the first bit. Then we have

$$\mathbb{E}\Phi(T_e f(X_1, x^n_2)) \leq \theta \mathbb{E}\Phi(T_e g_{j,k}(X_1, x^n_2)) + (1 - \theta) \mathbb{E}\Phi(T_e h_{j,k}(X_1, x^n_2)). \quad (15)$$

Similarly, we have

$$\mathbb{E}\Phi(T_e g_{j,k}(X_1, x^n_2)) = \mathbb{E}\Phi(T_e h_{j,k}(X_1, x^n_2)),$$

which follows from

$$T_e g_{j,k}(j, x^n_2) = T_e h(k, x^n_2).$$

$$T_e g_{j,k}(k, x^n_2) = T_e h(j, x^n_2).$$

and that $T_e g_{j,k}(x) = T_e h_{j,k}(x)$ for $x_1 \neq j, k$. Then inequality (15) becomes

$$\mathbb{E}\Phi(T_e f(X_1, x^n_2)) \leq \mathbb{E}\Phi(T_e g_{j,k}(X_1, x^n_2)).$$

We will have $\mathbb{E}\Phi(T_e f) \leq \mathbb{E}\Phi(T_e g_{j,k})$ by averaging over $x^n_1$. Repeat the argument for all such pairs $(j, k)$ and the last $n - 1$ coordinates. We will arrive at a monotone function. \qed
We have shown that our new definitions of the noise operator and the total influence are consistent with the old ones in the discrete cube case, i.e., \( p = 2 \). Also the total influence measures the changing rate of the moments of the noise operator. But the total influence lacks a geometric interpretation for higher values of \( p \).

This motivates the following modifications of the noise vector \( Z \) in Definition 5.1 and the random flipping operator \( \tilde{\sigma}_j \) (or \( \tilde{Z}_j \)) in Definition 5.2.

We will suppose that \( Z_i \) assigns 0 with probability \( 1 - \epsilon \) and \( \pm 1 \) (\( -1 = p - 1 \) in \( \mathbb{Z}/p\mathbb{Z} \)) with the same probability \( \epsilon/2 \). In other words, the noise can only change the alphabet in the transmission to its nearest values. Similar to (10), we have the following Fourier representation of \( T_\epsilon f \)

\[
T_\epsilon f(x) = \sum_\xi \prod_{j=1}^n [1 - \epsilon(1 - \cos(2\pi \xi_j/p))] \hat{f}(\xi)e_p(\xi \cdot x).
\]  

(16)

Accordingly, we let \( \tilde{Z}_j \) be a Bernoulli random variable taking 1 and \(-1\) with equal probability. Let \( S \) be the support of \( f \). We define \( j \)-th direction edge boundary

\[
\partial_j S = \{(x, y) : x_j - y_j \in \{\pm 1\}, x_k = y_k \text{ for } k \neq j\}
\]

and the edge boundary \( \partial S = \bigcup_{j=1}^n \partial_j S \). Analogies of identities (3) and (4) still hold. We have the following identities relating the edge boundary and the influence

\[
I_i(f) = \frac{|\partial_i S|}{p^n},
\]

(17)

\[
I(f) = \frac{|\partial S|}{p^n}.
\]

(18)

Similar to Proposition 5.3, we also have the following Fourier representation of the influence.

**Proposition 5.4.** Let \( f : (\mathbb{Z}/p\mathbb{Z})^n \to \{0, 1\} \) be a Boolean function. Then we have

\[
I_j(f) = 2 \sum_\xi (1 - \cos(2\pi \eta_j/p))|\hat{f}(\xi)|^2,
\]

and

\[
I(f) = 2 \sum_\xi \sum_{j=1}^n (1 - \cos(2\pi \eta_j/p))|\hat{f}(\xi)|^2.
\]

The following statement can be proved in the same manner as that of Theorem 3.1.

**Theorem 5.3.** Let \( \alpha > 1 \), \( n \) and \( \mathbb{E}f \) be fixed. When \( \epsilon \) is sufficiently small, \( \mathbb{E}(T_\epsilon f)^\alpha \) is maximized by a Boolean function with the least total influence, i.e., the Boolean function supported on the set of fixed size and least edge boundary.

**Remark 5.5.** Bollobás and Leader [2] proved sharp edge isoperimetric inequalities for the discrete torus and the grid (Theorem 8 and Theorem 3, respectively). When the subset possesses certain type of cardinalities, they know the extremal set; but, in general, they do not know which set to take, although they know the sharp bound of the edge boundary of the extremal sets.
Remark 5.6. Theorem 5.1 and Theorem 5.3 characterize the extremizers in a Fourier analytic way and a geometric way, respectively. This difference results from the two different definitions of the noise operator, which capture the group nature and the graph nature of the discrete torus, respectively.

Remark 5.7. We have the following analogy of Theorem 5.3 for general Markov semi-groups \((P_t)_{t \geq 0}\) defined by \((2)\). When \(t > 0\) is sufficiently small, \(\mathbb{E}(P_t f)^\alpha\) is maximized by the Boolean function supported on the set with the least edge boundary. This follows from the relation
\[
\frac{d}{dt} \mathbb{E}(P_t f)^\alpha \bigg|_{t=0} = -\alpha \mathbb{E}(L f) = -\frac{\alpha |\partial S|}{2 |E|},
\]
where \(L\) is the Laplacian, and \(|E|\) is the number of edges of the graph \(G\).

5.2 Tree

Now we discuss the problem of noise stability in a tree model, which was initially proposed by Mossel et al. [22] for the NICD problem.

We denote by \(T\) an undirected tree, which gives the geometry of the problem. The edges of \(T\) will be thought of as independent memoryless BSC(\(\epsilon\)) channels with the crossover probability \(\epsilon \in [0, 1/2]\). Let \(V\) denote the vertices of \(T\). We refer to \(S \subset V\) as the locations of the players. Some vertex \(u\) of \(T\) broadcasts a uniformly random string \(X_u \in \{0, 1\}^n\). This string follows the BSC edges of \(T\) and eventually reaches all vertices. It is easy to see that the choice \(u\) does not matter, in the sense that the resulting joint probability distribution on strings for all vertices is the same regardless of \(u\). Upon receiving their strings \(Y_v \in \{0, 1\}^n, v \in S\), each player applies a balanced Boolean function \(f_v : \{0, 1\}^n \mapsto \{0, 1\}\), producing one output bit. As usual, the goal of the players is to maximize
\[
\mathbb{E} \prod_{v \in S} f_v(Y_v) = \mathbb{P}(f_v(Y_v) = 1, v \in S)
\]
without any further communication. Note that the problem of \(\alpha\)-stability with \(\alpha = k\) considered in Section 3 is just this generalized noise stability on the star graph of \(k + 1\) vertices with the players at the \(k\) leaves.

In the case of NICD on the path graph, Mossel et al. [22] proved (Theorem 5.1) that the best strategy for all players is to use the same dictator function. In the general case, they showed (Theorem 6.3) that there always exists an optimal protocol in which all players use monotone functions. A careful check of their proofs shows that their arguments also yield the following analogs on the problem of noise stability. Hence, we omit the proofs.

**Theorem 5.4.** Suppose that \(T\) is a path of length \(k\) on the set \(\{0, 1, \ldots, k\}\). Let \(S = \{i_0, \ldots, i_l\}\) be a subset of size at least two. Then we have
\[
\mathbb{E} \prod_{v \in S} f_v(Y_v) \leq 2^{-l+1} \prod_{j=1}^{l} (1 + (1 - 2\epsilon)^{i_j - i_{j-1}}).
\]
Equality is achieved if and only if \(f_v\) are the identical dictator function.

**Theorem 5.5.** For any tree \(T\), the maximal correlation \(\mathbb{E} \prod_{v \in S} f_v(Y_v)\) can be achieved among monotone Boolean functions.
6 Discussion

We investigate the noise stability of Boolean functions in various settings, such as functions defined on the discrete cube, the discrete torus, as well as in a tree model. Characterizations of extremal functions are given in different scenarios. Close connections with the problem of non-interactive correlation distillation and Courtade-Kumar’s conjecture on the most informative Boolean function are discussed. This paper significantly generalizes our earlier work [20] with a focus on the discrete cube case. Regarding practical applications, our study of the discrete torus model is potentially useful for communications via low-noise channels with phase-shift keying (PSK) modulation [13, 12, 15, 23, 8]. For example, our study of the discrete torus model captures the character of the l-PSK schemes with errors limited to a phase shift of $\frac{2\pi}{l}$ or $-\frac{2\pi}{l}$, say each with probability $\epsilon/2$, i.e., the errors remain closest to the original signal. Future work may consider general non-negative functions on the discrete cube $\{0,1\}^n$ and Boolean functions on general product measure spaces. Analogous questions can be asked for general Markov semi-groups. Extension of the tree model in Section V to networks of general graphs is interesting from both theoretical and practical perspectives.

Acknowledgment

We would like to thank Alex Samorodnitsky for pointing out reference [1]. We are indebted to Imre Leader for his clarification of results in [2]. This work is supported by NSF grant CCF-1527270.

References

[1] V. Anantharam, A. Bogdanov, A. Chakrabarti, T. S. Jayram, and C. Nair. “A conjecture regarding optimality of the dictator function under Hellinger distance,” Available: http://chandra.ie.cuhk.edu.hk/pub/papers/HC/hel-conj.pdf

[2] B. Bollobás, and I. Leader. “Edge-isoperimetric inequalities in the grid,” Combinatorica., 11(4):299–314, 1991.

[3] N. H. Bshouty, J. C. Jackson, and C. Tamon. “Uniform-distribution attribute noise learnability,” Inform. and Comput., 187(2):277–290, 2003.

[4] I. Benjamini, G. Kalai, and O. Schramm. “Noise sensitivity of Boolean functions and applications to percolation,” Inst. Hautes Études Sci. Publ. Math., (90):5–43, 1999.

[5] C. Borell. “Geometric bounds on the Ornstein-Uhlenbeck velocity process,” Z. Wahrsch. Verw. Gebiete, 70(1):1–13, 1985.

[6] T. A. Courtade and G. R. Kumar. “Which Boolean functions maximize mutual information on noisy inputs?” IEEE Trans. Inform. Theory, 60(8):4515–4525, 2014.

[7] R. Eldan. “A two-sided estimate for the Gaussian noise stability deficit,” Invent. Math., 201(2):561–624, 2015.

[8] EM4100 Protocol description, Available at: http://www.priority1design.com.au/em4100_protocol.html

[9] L. H. Harper. “Optimal assignment of numbers to vertices,” SIAM J. Appl. Math., 12(1):131–135, 1964.
[10] J. Håstad. “Some optimal inapproximability results,” *J. ACM*, 48(4):798–859, 2001.

[11] W. Huleihel and O. Ordentlich. “How to quantize $n$ outputs of a binary symmetric channel to $n - 1$ bits?,” In *IEEE International Symposium on Information Theory*, 2017.

[12] IEEE Std 802.11b, Available: [https://www.standards.ieee.org/findstds/standard/802.11b-1999.html](https://www.standards.ieee.org/findstds/standard/802.11b-1999.html).

[13] IEEE Std 802.11g, Available: [https://www.standards.ieee.org/findstds/standard/802.11g-2003.html](https://www.standards.ieee.org/findstds/standard/802.11g-2003.html).

[14] M. Isaksson and E. Mossel. “Maximally stable Gaussian partitions with discrete applications,” *Israel J. Math.*, 189:347–396, 2012.

[15] S. Jayasimha and P. Jyothishendar, “Pragmatic tcm for 8-psk in satellite communications,” In *Proceedings of SPCOM*, pp. 129–134, Bangalore, India, June 2004.

[16] J. Kahn, G. Kalai, and N. Linial. “The influence of variables on boolean functions,” in *Proceedings of the 29th Annual IEEE Symposium on Fundations of Computer Science, IEEE Computer Society Press*, Los Alamitos, CA, pages 68–80, 1988.

[17] D. J. Kleitman. “Families of non-disjoint subsets,” *J. Combinatorial Theory*, 1:153–155, 1966.

[18] A. R. Klivans, R. O’Donnell, and R. A. Servedio. “Learning intersections and thresholds of halfspaces,” *J. Comput. System Sci.*, 68(4):808–840, 2004.

[19] G. Kindler, R. O’Donnell, and D. Witmer. “Remarks on the most informative function conjecture at fixed mean,” Available: [https://arxiv.org/abs/1506.03167](https://arxiv.org/abs/1506.03167).

[20] J. Li, and M. Médard. “Boolean functions: noise stability, non-interactive correlation, and mutual information,” In *Proc. IEEE Intl. Symp. Inform. Theory*, Vale, USA, July 2018.

[21] E. Mossel and R. O’Donnell. “Coin flipping from a cosmic source: on error correction of truly random bits,” *Random Structures Algorithms*, 26(4):418–436, 2005.

[22] E. Mossel, R. O’Donnell, O. Regev, J. E. Steif, and B. Sudakov. “Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami-Beckner inequality,” *Israel J. Math.*, 154:299–336, 2006.

[23] National Instruments, “Advanced rfid measurements: Basic theory to protocol conformance test,” Available: [http://www.ni.com/tutorial/6645/en/](http://www.ni.com/tutorial/6645/en/), Dec 19, 2013.

[24] R. O’Donnell. “Hardness amplification within NP,” *J. Comput. System Sci.*, 69(1):68–94, 2004.

[25] R. O’Donnell. “Analysis of Boolean Functions,” Cambridge University Press, June 2014.

[26] O. Ordentlich, O. Shayevitz, and O. Weinstein. “Dictatorship is the most informative balanced function at the extremes,” In *Electronic Colloquium on Computational Complexity*, No. 84, 2015.

[27] G. Pichler, P. Piantanida, and Gerald Matz. “Dictator functions maximize mutual information,” Available: [https://arxiv.org/abs/1604.02109](https://arxiv.org/abs/1604.02109).
[28] A. Samorodnitsky. “On the entropy of a noisy function,” *IEEE Trans. Inform. Theory*, 62(10):5446–5464, 2016.

[29] K. Yang. “On the (im)possibility of non-interactive correlation distillation,” *Theoret. Comput. Sci.*, 382(2):157–166, 2007.