SELF-PREDICTING BOOLEAN FUNCTIONS

NIR WEINBERGER AND OFER SHAYEVITZ

ABSTRACT. A Boolean function g is said to be an *optimal predictor* for another Boolean function f, if it minimizes the probability that $f(X^n) \neq g(Y^n)$ among all functions, where X^n is uniform over the Hamming cube and Y^n is obtained from X^n by independently flipping each coordinate with probability δ. This paper is about *self-predicting functions*, which are those that coincide with their optimal predictor.

1. INTRODUCTION

One of the most important properties of a Boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$ is its robustness to noise in its inputs. This robustness is traditionally measured by the *noise sensitivity* of the function

$$\text{NS}_\delta[f] := \Pr\left(f(X^n) \neq f(Y^n)\right), \tag{1}$$

where $X^n \in \{-1, 1\}^n$ is a uniform Bernoulli vector, and $Y^n \in \{-1, 1\}^n$ is obtained from X^n be flipping each coordinate independently with probability $0 < \delta < 1/2$. The noise sensitivity of Boolean functions has been extensively investigated [O'D14], most often in terms of the equivalent notion of *stability*

$$\text{Stab}_\rho[f] := \mathbb{E}\left[f(X^n)f(Y^n)\right],$$

where $0 < \rho < 1$ is the correlation parameter, i.e., $\rho := \mathbb{E}(X_i Y_i) = 1 - 2\delta$. The noise sensitivity of f can also be interpreted as the error probability of a predictor trying to guess the value of $f(X^n)$ by simply applying f to the noisy input Y^n. While this predictor is intuitively appealing and easy to analyze, it is generally suboptimal. As a simple example, think of the case where f is biased and the noise level δ is sufficiently high; it is easy to see that a constant predictor would result in a lower error probability than $f(Y^n)$ would.

The optimal predictor, i.e., the one that minimizes the error probability in predicting $f(X^n)$ from Y^n, is clearly given by the sign of $\mathbb{E}(f(X^n) \mid Y^n = y^n)$. In general, this

function might be rather different from f itself. However, while using the optimal predictor is generally superior to using the function itself (albeit as we shall see, by a factor of two at the most), computing the former is often very difficult as its value in an point depends on the values of the function over the entire Hamming cube. It is therefore interesting to study functions that coincide with their optimal predictor; we call these functions *self-predicting* (SP).

Clearly, SP functions exhibit a desirable property - the optimal prediction of the function is obtained by simply applying it to the noisy inputs. For example, suppose the function describes a voting rule and the noise represents possible contamination of the votes (e.g., due to fraud). If the function is SP, then any mechanism used for computing the function with clean votes can be used without any modification in case it turns out that the votes are actually noisy. In this case, the output of the function is the optimal predictor for its true value. It should be noted, however, that being SP does not imply anything about the ordinary stability of the function. For example, all parity functions (characters) are SP functions, including the least stable one, to wit, the parity of all inputs (namely, the largest character). Nonetheless, if, e.g., there are a few alternatives for choosing a function to be used, and all of these functions have the same stability, it is sensible to choose one of the SP functions among them (if such exists).

Nonetheless, a function can be SP at certain noise levels but not at others. We thus say that a function is *uniformly SP (USP)* if it is SP at any noise level. For example, in the voting scenario mentioned above, it may not be realistic to assume that the noise level is known, yet if the function is USP it can always be used to obtain the optimal prediction of the true voting result.

In this paper, we introduce and explore self-predictability of Boolean functions. We derive various properties of SP functions, and specifically the following:

- If a function is monotone (resp. odd, resp. symmetric), then so is the optimal predictor. We use this fact to show that Majority functions are USP, and that for a monotone function, self-predictability at dominating boundary points is necessary and sufficient for a function to be SP.
- SP at high correlation: A function with Fourier degree k is SP for any $\rho > 1 - 1/k^2$, and if f is SP for $\rho > 1 - \varepsilon$ and $n = \Omega(1/\varepsilon)$, then each point x^n has a distance-2 neighbor with the same function value.
- SP at low correlation: Any function for which there exists ρ^* such that it is SP for all $\rho \in [0, \rho^*]$ (abbreviated LCSP) is spectral threshold, i.e., equal to the sign of its lowest Fourier level. This simple fact implies many properties: LCSP functions are either balanced or constant, they have energy at least $1/2$

on their first level (if any), and a monotone LCSP function is $\sqrt{\frac{2}{\pi n}}$-close to a linear threshold function.

- Sharp threshold: All functions are trivially SP for $\rho > 1 - \frac{2 \ln 2}{n} + O(n^{-2})$. However, the fraction of SP functions is doubly-exponential small with n whenever the correlation parameter is either $\rho = 1 - \frac{2\alpha}{n}$ for $\alpha > 1$ or $\rho = 1 - 2\delta$ for $\delta \in [0, \delta_{\max}]$, $\delta_{\max} \approx 0.097$.

The paper is organized as follows. Section 2 contains basic notation and Fourier-theory facts. The self-predictability problem and some basic properties are introduced in Section 3, including the proof that Majority is USP. Section 4 discusses high-correlation sufficient conditions for SP, and Section 5 discusses low-correlation SP functions. Section 6 provides stability-based necessary conditions for SP. In Section 7, a *sharp threshold* phenomenon is proved for the SP property. The paper is concluded in Section 8 with a list of open problems.

2. PRELIMINARIES

2.1. **Notation and Definitions.** We use upper case letters for random variables and random vectors, and their lower case counterparts for specific realizations. For vectors we write $x_i^j = (x_i, \ldots, x_j)$ and omit the subscript whenever $i = 1$, and denote a concatenation of vectors by $(x_i^j, x_k^m) = (x_i, \ldots, x_j, x_k, \ldots, x_m)$. We denote the cardinality of a set S by $|S|$, the complement of the set $\mathcal{A}$ by $\mathcal{A}^c$, and write $[n]$ for the set $\{1, 2, \ldots, n\}$. We define the indicator function by $\mathbb{1}(\cdot)$, the sign function by $\text{sgn}(z)$ where by convention $\text{sgn}(0) = 0$, unless otherwise stated. Throughout, the logarithm $\log(t)$ is base 2, while $\ln(t)$ is the natural logarithm. The Hamming distance between x^n and y^n is $d_H(x^n, y^n)$.

In this paper, X^n is a uniformly distributed binary vector, and Y^n is the binary vector obtained by flipping each coordinate of X^n with some given probability $\delta \in [0, 1/2]$. We write $p(x^n, y^n)$ to denote the associated joint probability mass function, and $p(x^n \mid y^n)$, e.g., to denote the conditional probability mass function. As a binary alphabet, for the most part we will find it convenient to work with $\{-1, 1\}$, in which case it is more natural to consider the *correlation* parameter $\rho := \mathbb{E}(X_i Y_i) = 1 - 2\delta \in [0, 1]$ instead of the crossover probability parameter δ. We will use the latter notations throughout the paper, with the exception of a few proofs where we find it more convenient to work with either δ or the binary alphabet $\{0, 1\}$.

2.2. **Boolean Functions and Fourier Analysis.** In this paper we consider Boolean functions $f : \{-1, 1\}^n \to \{-1, 1\}$. The *distance* between two Boolean functions f and g is defined as the fraction of inputs on which they disagree, i.e., $\Pr(f(X^n) \neq g(X^n))$. We say that f and g are ε-*close* if their distance is at most ε.

An inner product between two Boolean functions f, g is defined as

$$(2) \qquad \langle f, g \rangle := \mathbb{E}\left(f(X^n)g(X^n)\right).$$

A *character* associated with a set of coordinates $S \subseteq [n]$ is the Boolean function $x^S := \prod_{i \in S} x_i$, where by convention $x^\emptyset = 1$. It can be shown [O'D14, Chapter 1] that the set of all characters form an orthonormal basis with respect to (w.r.t.) the inner product (2). Furthermore,

$$f(x^n) = \sum_{S \subseteq [n]} \hat{f}_S \cdot x^S,$$

where $\{\hat{f}_S\}_{S \subseteq [n]}$ are the *Fourier coefficients* of f, given by $\hat{f}_S = \langle x^S, f \rangle = \mathbb{E}(X^S \cdot f(X^n))$. When S is a singleton $\{i\} \subset [n]$, we use the shorthand $\hat{f}_i = \hat{f}_{\{i\}}$. The *Fourier weight* of f at degree k is

$$W^k[f] := \sum_{S \subseteq [n] : |S| = k} \hat{f}_S^2.$$

Instead of the *noise sensitivity* defined in (1) it is more common to consider the *stability*, defined as

$$\mathrm{Stab}_\rho[f] := \mathbb{E}\left(f(X^n)f(Y^n)\right),$$

where the noise sensitivity and stability are trivially related via

$$\mathrm{Stab}_\rho[f] = 1 - 2\,\mathrm{NS}_{\frac{1-\rho}{2}}[f].$$

Thus, the stability of a function is directly related to the error probability of the possibly suboptimal predictor $f(y^n)$ to the function's true value $f(x^n)$.

The *noise operator* for ρ-correlated X^n and Y^n is defined as

$$T_\rho f(y^n) := \mathbb{E}\left(f(X^n) \mid Y^n = y^n\right),$$

and, evidently, as $\{(X_i, Y_i)\}$ is an i.i.d. sequence,

$$(3) \qquad T_\rho f(y^n) = \mathbb{E}\left(\sum_{S \subseteq [n]} \hat{f}_S \cdot X^S \mid Y^n = y^n\right) = \sum_{S \subseteq [n]} \rho^{|S|} \cdot \hat{f}_S \cdot y^S.$$

The stability can then be expressed using the Fourier coefficients and the noise operator as

$$\mathrm{Stab}_\rho[f] = \mathbb{E}\left(\mathbb{E}\left(f(X^n)f(Y^n)\right) \mid Y^n\right)$$
$$= \mathbb{E}\left(f(Y^n)T_\rho f(Y^n)\right)$$
$$= \langle f, T_\rho f \rangle$$
$$(4) \qquad = \sum_{S \subseteq [n]} \rho^{|S|} \cdot \hat{f}_S^2$$

$$= \left\| T_{\sqrt{\rho}} f \right\|_2^2,$$

where (4) follows from *Plancherel's identity* $\langle f, g \rangle = \mathbb{E}(f(X^n)g(X^n)) = \sum_{S \subseteq [n]} \hat{f}_S \hat{g}_S$.

A Boolean function f is called a *linear threshold function (LTF)* if there exists coefficients $a_0^n \in \mathbb{R}^{n+1}$ such that

$$f(x^n) = \mathrm{sgn} \left(a_0 + \sum_{i=1}^n a_i x_i \right).$$

Note that if $a_0 = 0$ then f is *balanced*, i.e., $\Pr(f(X^n) = 1) = 1/2$. More generally, a function f is a *polynomial threshold function* (PTF) [Bru90] of degree k if there exists $\{\hat{p}_S\}$ such that $\max_{S:\hat{p}_S \neq 0} |S| = k$ and

(5)
$$f(x^n) = \mathrm{sgn} \left(\sum_{S \subseteq [n]} \hat{p}_S \cdot x^S \right).$$

A PTF has *sparsity* s if $\{\hat{p}_S\}$ is supported over exactly s terms. For LTF and PTFs, we will always assume that coefficients are chosen such that the polynomial inside the sign operator is never identically zero.

3. Optimal Prediction and Self Predicting (SP) Functions

3.1. The Optimal Predictor.
Let $f : \{-1, 1\}^n \to \{-1, 1\}$ be some Boolean function. It is easy to see that the optimal predictor (minimizing the error probability) of $f(X^n)$ given that $Y^n = y^n$ has been observed, is simply

$$\mathrm{sgn}\, \mathbb{E}\left(f(X^n) \mid Y^n = y^n \right) = \mathrm{sgn}\, T_\rho f(y^n).$$

Note that according to our definition $\mathrm{sgn}(0) = 0$, but ties can of course be broken arbitrarily in any other way.

The optimal predictor preserves several properties of the function. We define the natural partial order $\preceq$ over $\mathbb{R}^k$, where $y^k \preceq z^k$ if and only if $y_i \leq z_i$ for all coordinates i. We write $\prec$ to denote the case of strict inequality in at least one of the coordinates.

Recall that [O'D14, Definition 2.8.] a function $f : \{-1, 1\}^n \to \mathbb{R}$ is called:

- *Monotone on $S \subseteq [n]$*, if $f(y^n) \leq f(z^n)$ whenever both $y^S \preceq z^S$ and $y^{[n] \setminus S} = z^{[n] \setminus S}$, and *monotone* if it is monotone on $[n]$.
- *Odd (resp. even) if $f(x^n) = -f(-x^n)$ for all $x^n \in \{-1, 1\}^n$ (resp. $f(x^n) = f(-x^n)$).
- *Symmetric* if $f(\pi(x^n)) = f(x^n)$ for all $x^n \in \{-1, 1\}^n$ and permutation $\pi \in \mathbb{S}_n$ (where $\mathbb{S}_n$ is the symmetric group over the set $[n]$) and $\pi(x^n) = (x_{\pi(1)}, \ldots, x_{\pi(n)})$.

Proposition 3.1. *For $\rho \in (0, 1]$, $\mathrm{sgn}\, T_\rho(\cdot)$ preserves monotonicity on any $S \subseteq [n]$, parity (oddness or evenness) and symmetry.*

Proof.

- *Monotonicity*: This property stems from the fact that the operator T_ρ itself preserves monotonicity (for $\rho \in (0,1]$) [Kel10, Proof of Proposition 4.4], [KKM16, Claim 2.4. (b)]. A short proof is given for the sake of completeness. Assume that $f(y^n) = 1$ and let z^n satisfy $y^S \preceq z^S$ and $y^{[n]\setminus S} = z^{[n]\setminus S}$. We prove the statement for a singleton S, say $S = \{n\}$. The general case then follows by applying the same argument repeatedly. If $y_n = 1$ the claim is trivial. Assume $y_n = -1$ and let z^n agree with y^n except on the nth coordinate. Due to monotonicity of f, we have that $f(z^n) = 1$. Then

$$
\begin{aligned}
T_\rho f(z^n) &= \sum_{x^n} p(x^n \mid z^n) f(x^n) \\
&= \sum_{x^{n-1}} \sum_{x_n} p(x^{n-1} \mid y^{n-1}) p(x_n \mid 1) f(x^n) \\
&= \sum_{x^{n-1}} p(x^{n-1} \mid y^{n-1}) \left[\delta f(x^{n-1}, -1) + (1 - \delta) f(x^{n-1}, 1) \right] \\
&\geq \sum_{x^{n-1}} p(x^{n-1} \mid y^{n-1}) \left[(1 - \delta) f(x^{n-1}, -1) + \delta f(x^{n-1}, 1) \right] \\
&= T_\rho f(y^n)
\end{aligned}
$$

  where the inequality holds since f is monotone on the nth coordinate (and $\delta \in [0, 1/2)$). Hence, $\operatorname{sgn} T_\rho f(z^n) \geq \operatorname{sgn} T_\rho f(y^n)$.

- *Parity*: f is odd if and only if $\hat{f}_S = 0$ for all $S \subseteq [n]$ such that $|S|$ is even [O'D14, Exercise 1.8]. It follows from the Fourier expansion of $T_\rho f$ (3) that if f is odd then so is $T_\rho f$, i.e. $T_\rho f(x^n) + T_\rho f(-x^n) = 0$ for all $x^n \in \{-1, 1\}^n$. Thus, $\operatorname{sgn} T_\rho f$ is also odd (utilizing the convention $\operatorname{sgn}(0) = 0$). The proof for even functions is similar.

- *Symmetry*: f is symmetric if and only if $\hat{f}_S$ depends on S only via $|S|$. Hence (3) implies that if f is symmetric then so is $T_\rho f$. A composition of scalar function and a symmetric function results in a symmetric function, and thus $\operatorname{sgn} T_\rho f$ is symmetric.

$\square$

We say that a Boolean function f is *ρ-self-predicting (ρ-SP) at y^n*, if the optimal predictor given y^n at correlation level ρ coincides with the function itself whenever it is not tied, i.e., if

$$
f(y^n) = \operatorname{sgn} T_\rho f(y^n),
$$

whenever $T_\rho f(y^n) \neq 0$. The function f is called *ρ-SP* if it is ρ-SP for any $y^n \in \{-1, 1\}^n$. We say that f is *uniformly self-predicting (USP)* if it is ρ-SP for any $\rho \in [0, 1]$. We also

say that f is *low-correlation self-predicting (LCSP)*, if there exists some $\rho^* > 0$ such that f is ρ-SP for all $\rho \in [0, \rho^*)$.

We note in passing that seemingly plausible properties may not hold in general:

Example 3.2. The optimal predictor of a balanced function may not be balanced. For example, the function

$$\frac{1}{4}(2x_1 + x_3 - 2x_1x_2 + x_1x_3 + x_2x_3 - x_3x_4$$
$$+ x_1x_2x_3 + x_1x_3x_4 - x_2x_3x_4 + x_1x_2x_3x_4)$$

is a balanced function, yet $\operatorname{sgn} T_\rho f$ is non-balanced when $\rho = 1/2$.

Example 3.3. In the following sections we explore functions that are SP for high or low correlation. However, self-predictability is not necessarily a monotone property in ρ. To wit, if a function is ρ_0-SP then it might not be ρ-SP for some $\rho \geq \rho_0$. Indeed, there are functions that admit such an "irregular" behavior. We have numerically analyzed LTFs with randomly drawn coefficients, and found, for example, that the balanced LTF with $n = 11$ and coefficients

$$a_1^{11} = (13, 43, 67, 67, 67, 117, 153, 165, 165, 179, 179)$$

is ρ-SP only for $\rho \in [0, 0.312] \cup (0.544, 1]$.

3.2. **Elementary USP Functions.** The following fact follows easily from the definition.

Proposition 3.4. *All the characters are USP.*

Proof. Let $f(x^n) = x^S$ for some $S \subseteq [n]$. Then for any y^n,

$$\operatorname{sgn} T_\rho f(y^n) = \operatorname{sgn}\left(\rho^{|S|} \cdot y^S\right)$$
$$= \operatorname{sgn}\left(y^S\right)$$
$$= f(y^n).$$

$\square$

We next show that Majority (for odd n), given by,

$$\operatorname{Maj}(x^n) := \operatorname{sgn} \sum_{i \in [n]} x_i$$

is USP. While this property is plausible, it does not stem from only analyzing the "local" behavior of the function. Specifically, at a boundary point y^n, i.e., one for which $\sum_{i \in [n]} y_i = \pm 1$, there are more neighbors in the immediate neighborhood of y^n (say, Hamming distance one or two) who disagree with y^n on the value of the function, than

those who agree with it. Thus, any proof that such a point is SP for all $\rho \in (0, 1]$ cannot rely only on the local values of the function in the vicinity of that point. Rather, it should take into account the function's value in larger neighborhoods, or even over the entire Hamming cube.

Theorem 3.5. *Majority is USP.*

Proof. Since Maj is monotone, odd and symmetric, then so is $\mathrm{sgn}\, T_\rho \mathrm{Maj}$ (Proposition 3.1). Hence, for all $x^n \in \{-1, 1\}^n$

$$\mathrm{sgn}\, T_\rho \mathrm{Maj}(x^n) + \mathrm{sgn}\, T_\rho \mathrm{Maj}(-x^n) = 0. \tag{6}$$

Consider without loss of generality x^n such that $\mathrm{Maj}(x^n) = 1$, i.e., if w is the number of 1's in x^n, then $w > n - w$. Then, $\overline{x} = (1^w, -1^{n-w})$ and $\tilde{x} = (1^{n-w}, -1^w)$ satisfy $\tilde{x}^n \preceq \overline{x}^n$, and from symmetry,

$$\mathrm{sgn}\, T_\rho \mathrm{Maj}(x^n) = \mathrm{sgn}\, T_\rho \mathrm{Maj}(\overline{x}^n) \geq \mathrm{sgn}\, T_\rho \mathrm{Maj}(\tilde{x}^n) = \mathrm{sgn}\, T_\rho \mathrm{Maj}(-x^n).$$

Hence, (6) implies that $\mathrm{sgn}\, T_\rho \mathrm{Maj}(x^n) \geq 0$, as was required to be proved. $\square$

Remark 3.6. An indirect way of proving Theorem 3.5 is via May's theorem [O'D14, Ex. 2.3]: Since $\mathrm{sgn}\, T_\rho \mathrm{Maj}$ is monotone, odd and symmetric, it must be the majority function itself.

By numerically experimenting with simple LTFs one can find that Majority (and characters) are not the only USP functions, and not even the only USP LTFs. Specifically:

Example 3.7. The balanced LTFs with $n = 5$ and coefficients $a_1^5 = (1, 1, 3, 3, 5)$, with $n = 7$ and coefficients $a_1^7 = (1, 1, 3, 3, 3, 5, 7)$, with $n = 9$ and coefficients $a_1^9 = (1, 1, 3, 3, 3, 5, 5, 5, 7)$, with $n = 11$ and coefficients $a_1^{11} = (1, 1, 3, 3, 3, 3, 5, 5, 5, 7, 7)$ can all be verified by direct computation to be USP.

In the next section we generate classes of USP functions by utilizing operations which preserve the SP property.

3.3. **SP/USP Preserving Operators.** Let us next discuss several operations that preserve self-predictability. First, we note that self-predictability is invariant to negation of inputs. We write $\circ$ for the Hadamard product.

Proposition 3.8. *Let $a^n \in \{-1, 1\}^n$. Then, $f(x^n)$ is ρ-SP if and only if $f(a^n \circ x^n)$ is ρ-SP.*

The straightforward proof is omitted. Next, we consider the case of separable functions.

Proposition 3.9. *Let $f(x^n) = g(x_1^k) \cdot h(x_{k+1}^n)$. Then f is ρ-SP if and only if both g and h are ρ-SP.*

Proof. If g and h are both ρ-SP then for any y^n,

$$\begin{aligned}
\operatorname{sgn} T_\rho f(y^n) &= \operatorname{sgn} T_\rho \left(g(y^k) \cdot h(y_{k+1}^n) \right) \\
&= \operatorname{sgn} \left(T_\rho g(y^k) \cdot T_\rho h(y_{k+1}^n) \right) \\
&= g(y_1^k) \cdot h(y_{k+1}^n) \\
&= f(y^n).
\end{aligned}$$

Conversely, suppose that f is ρ-SP for some $\rho \in (0, 1]$. Then there must exist at least one point y_{k+1}^n at which h is ρ-SP, since if this was not the case, then $T_\rho h(y_{k+1}^n) \cdot f(h_{k+1}^n) \le 0$ holds for all h_{k+1}^n. This, however, is impossible since

$$\mathbb{E}\left[T_\rho h(Y_{k+1}^n) \cdot h(Y_{k+1}^n) \right] = \sum \rho^{|S|} \hat{h}_S^2 > 0.$$

Hence, without loss of generality, we may assume that $h(y_{k+1}^n) = 1$. Then for any y^k

$$\begin{aligned}
\operatorname{sgn} T_\rho g(y^k) &= \operatorname{sgn} T_\rho g(y^k) \cdot \operatorname{sgn} T_\rho h(y_{k+1}^n) \\
&= \operatorname{sgn} \left(T_\rho g(y^k) \cdot T_\rho h(y_{k+1}^n) \right) \\
&= \operatorname{sgn} T_\rho f(y^n) \\
&= f(y^n) \\
&= g(y^k) \cdot h(y_{k+1}^n) \\
&= g(y^k).
\end{aligned}$$

Hence g, and symmetrically, also h, are ρ-SP. $\qquad\square$

Note that Proposition 3.4 also follows as a simple corollary to Proposition 3.9. Next, we consider functions of equal-size disjoint characters.

Proposition 3.10. *Let $\{S_\ell \subseteq [n]\}_{\ell \in [m]}$ be disjoint subsets of equal size $|S_\ell| = w$. Let $f : \{-1, 1\}^m \to \{-1, 1\}$ be ρ^w-SP. Then $f(x^{S_1}, x^{S_2}, \dots, x^{S_m})$ is ρ-SP.*

Proof. By equating coefficients of the Fourier representation (which are unique), it is readily obtained that the Fourier coefficients of $h(x^n) = f(x^{S_1}, x^{S_2}, \dots, x^{S_m})$ are given by

$$\hat{h}_S = \begin{cases} \hat{f}_T, & S = \cup_{t \in T} S_t \\ 0, & \text{otherwise} \end{cases}.$$

Hence,

$$\operatorname{sgn} T_\rho h(y^n) = \operatorname{sgn} \sum_{S \subseteq [n]} \rho^{|S|} \hat{h}_S y^S$$

$$= \text{sgn} \sum_{T \subseteq [m]} \rho^{w|T|} \cdot \hat{h}_{\cup_{t \in T} S_t} \cdot y^{\cup_{t \in T} S_t}$$

$$= \text{sgn} \sum_{T \subseteq [m]} \rho^{w|T|} \hat{f}_T \prod_{t \in T} y^{S_t}$$

$$= \text{sgn} \, T_{\rho^w} f(y^{S_1}, y^{S_2}, \ldots, y^{S_m})$$

$$= f(y^{S_1}, y^{S_2}, \ldots, y^{S_m})$$

$$= h(y^n).$$

$\square$

Example 3.11. Using the fact that characters and Majority are USP functions, together with Propositions 3.8, 3.9 and 3.10, we can construct many distinct USP functions. For example, the function

$$\text{sgn} \left((x_1 x_2 + x_3 x_4 + x_5 x_6) \cdot (x_7 x_8 x_9 - x_{10} x_{11} x_{12} - x_{13} x_{14} x_{15}) \cdot x_{16} \right)$$

is USP.

Nonetheless, there are USP functions that cannot be constructed from characters and Majority this way. For example, none of these functions can be an LTF, as the USP functions in Example 3.7.

3.4. **Closeness to SP and Strong Stability.** How far can a function be from self predicting? We say that a function is ε-*close to ρ-SP*, to mean that f and its optimal predictor $\text{sgn} \, T_\rho f$ are ε-close.

Lemma 3.12. *Any function f is $\sum_{S \subseteq [n]} (1 - \rho^{|S|}) \hat{f}_S^2$-close to ρ-SP.*

Proof. Let $A \subseteq \{-1, 1\}^n$ be the set of all y^n at which f is ρ-SP. Hence for any $y^n \notin A$ it must be that $f(y^n) \cdot T_\rho f(y^n) < 0$. Noting that $|T_\rho f(y^n)| \leq 1$, we have that

$$\mathbb{E} \left(f(Y^n) \cdot T_\rho f(Y^n) \right) \leq \Pr \left(Y^n \in A \right).$$

On the other hand, it also holds that

$$\mathbb{E} \left(f(Y^n) \cdot T_\rho f(Y^n) \right) = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}_S^2.$$

The proof now follows by recalling that $\sum_{S \subseteq [n]} \hat{f}_S^2 = 1$. $\square$

For any n, functions that depend on all n variables can be found (even balanced ones), whose distance from their optimal predictor is larger than some universal constant. The problem with this measure of closeness to SP is that in many cases the optimal predictor might be different from the functions on inputs that are very noisy, i.e., where the posterior probability of the function value is close to uniform. Thus, a more practically

motivated way of quantifying closeness to SP is by considering noise sensitivity and stability.

Define the *strong noise sensitivity* of a function f to be

$$\mathrm{NS}_\delta^*[f] := \Pr\left(f(X^n) \neq \operatorname{sgn} T_\rho f(Y^n)\right),$$

and the associated *strong stability* as

$$\mathrm{Stab}_\rho^*[f] := \mathbb{E}\left(f(X^n) \cdot \operatorname{sgn} T_\rho f(Y^n)\right).$$

Of course, just as for the regular noise sensitivity and stability, we have the trivial connection

$$\mathrm{Stab}_\rho^*[f] = 1 - 2\,\mathrm{NS}_{\frac{1-\rho}{2}}^*[f].$$

We can also express the strong stability in terms of the noise operator:

$$
\begin{aligned}
\mathrm{Stab}_\rho^*[f] &= \mathbb{E}\left(\mathbb{E}\left(f(X^n) \cdot \operatorname{sgn} T_\rho f(Y^n) \mid Y^n\right)\right) \\
&= \mathbb{E}\left(T_\rho f(Y^n) \cdot \operatorname{sgn} T_\rho f(Y^n)\right) \\
&= \mathbb{E}\left|T_\rho f(Y^n)\right| \\
&= \|T_\rho f\|_1.
\end{aligned}
$$

Thus the 1-norm of $T_\rho f$ can be interpreted in terms of the error probability associated with the optimal predictor for f. Since the optimal predictor $\operatorname{sgn} T_\rho f$ can only do better than f itself, we immediately have:

Proposition 3.13. *For any function f and any $\rho \in [0,1]$*

$$\left\|T_{\sqrt{\rho}} f\right\|_2^2 \leq \|T_\rho f\|_1,$$

with equality if and only if f is ρ-SP.

The strong stability can also be upper bounded by a regular stability expression.

Proposition 3.14. $\mathrm{Stab}_\rho[f] \leq \mathrm{Stab}_\rho^*[f] \leq \sqrt{\mathrm{Stab}_{\rho^2}[f]}.$

Proof. Write

$$
\begin{aligned}
\mathrm{Stab}_\rho^*[f] &= \langle T_\rho f, \operatorname{sgn} T_\rho f \rangle \\
&\leq \|T_\rho f\|_2 \cdot \|\operatorname{sgn} T_\rho f\|_2 \\
&= \sqrt{\langle T_\rho f, T_\rho f \rangle} \\
&= \sqrt{\langle T_{\rho^2} f, f \rangle} \\
&= \sqrt{\mathrm{Stab}_{\rho^2}[f]}.
\end{aligned}
$$

where the inequality is by Cauchy-Schwartz's inequality, the next equality is since $\|\operatorname{sgn} T_\rho f\|_2 = 1$, and the following equality is since $T_\rho f$ is a self-adjoint operator (this follows from Plancherel's identity: $\langle T_\rho f, g\rangle = \sum_{S\subseteq[n]}\rho^{|S|}\hat{f}_S\hat{g}_S = \langle f, T_\rho g\rangle$). $\qquad\square$

An immediate consequence of the above is:

Corollary 3.15. *The strong noise sensitivity satisfies:*

$$\frac{1 - \sqrt{\operatorname{Stab}_{\rho^2}[f]}}{1 - \operatorname{Stab}_\rho[f]} \cdot \operatorname{NS}_\delta[f] \le \operatorname{NS}_\delta^*[f] \le \operatorname{NS}_\delta[f].$$

Note that this bound is tight for the characters (and again shows that they are USP). We can easily derive the following weaker statements:

Corollary 3.16. *For any f*

(7)
$$\frac{\operatorname{NS}_\delta[f]}{2} \le \operatorname{NS}_\delta^*[f] \le \operatorname{NS}_\delta[f].$$

If f is balanced, then

(8)
$$\frac{\operatorname{NS}_\delta[f]}{1 + \rho} \le \operatorname{NS}_\delta^*[f] \le \operatorname{NS}_\delta[f].$$

Proof. The bounds in (7) follow from $\operatorname{Stab}_{\rho^2}[f] \le \operatorname{Stab}_\rho[f]$ and $\min_{t\in[0,1]}\frac{1-\sqrt{t}}{1-t} = 1/2$. The bounds in (8) follow from

$$\frac{1 - \sqrt{\operatorname{Stab}_{\rho^2}[f]}}{1 - \operatorname{Stab}_\rho[f]} \ge \frac{1 - \sqrt{\operatorname{Stab}_{\rho^2}[f]}}{1 - \operatorname{Stab}_{\rho^2}[f]}$$

$$= \frac{1}{1 + \sqrt{\operatorname{Stab}_{\rho^2}[f]}}$$

$$\ge \frac{1}{1 + \rho}.$$

$\qquad\square$

We may obtain improved bounds for low correlation values:

Proposition 3.17. *Suppose $W^1[f] > 0$. Then:*

$$\max\left\{1, \frac{1}{\sqrt{2W^1[f]}} + O(\rho^2)\right\} \le \frac{\operatorname{Stab}_\rho^*[f]}{\operatorname{Stab}_\rho[f]} \le \frac{1}{\sqrt{W^1[f]}} + O(\rho^2).$$

Proof. We have that

$$\operatorname{Stab}_\rho^*[f] = \mathbb{E}\,|T_\rho f(Y^n)|$$

$$= \mathbb{E}\left|\sum_{i=1}^n \rho\hat{f}_i Y_i\right| + O(\rho^2).$$

Khintchine's inequality [Haa81] then implies

$$\frac{1}{\sqrt{2}} \cdot \sqrt{W^1[f]} \cdot \rho + O(\rho^2) \leq \mathrm{Stab}^*_\rho[f] \leq \sqrt{W^1[f]} \cdot \rho + O(\rho^2),$$

and the result follows from [O'D14, Proposition 2.51]

$$\mathrm{Stab}_\rho[f] = W^1[f] \cdot \rho + O(\rho^2).$$

□

Corollary 3.18. *For any balanced LTF $W^1[f] \geq 1/2$ [O'D14, Theorem 5.2], and so*

$$\frac{\mathrm{Stab}^*_\rho[f]}{\mathrm{Stab}_\rho[f]} \leq \sqrt{2} + O(\rho^2).$$

4. HIGH CORRELATION SUFFICIENT CONDITIONS

In this section, we derive sufficient conditions on a function to be SP using various arguments. All our conditions will be high correlation ones, i.e., for ρ_0 larger than some threshold. To that end, we will need a simple characterization of monotone SP functions. Recall that x^n is called a *boundary point* of f if the value of $f(x^n)$ can be flipped by flipping some single coordinate of x^n. We further say that x^n is a *dominating boundary point* of f if $f(x^n) = 1$ (resp. $= -1$) and $f(y^n) = -1$ (resp. $= 1$) for any $y^n \prec x^n$ (resp. $x^n \prec y^n$).

The following is a simple corollary to the fact that monotonicity is preserved by $\mathrm{sgn}\, T_\rho$ (Proposition 3.1).

Proposition 4.1. *A monotone function is ρ-SP if and only if it is ρ-SP at all its dominating boundary points.*

We can now prove the following:

Proposition 4.2. *Any function is ρ-SP for $\rho > 2^{(n-1)/n} - 1$, and there is no better universal guarantee.*

Proof. This range corresponds to the values of the crossover probability $\delta \in [0, 1-2^{-1/n})$ for which the probability no bit was flipped $(1 - \delta)^n$, is at least $1/2$. This bound is achieved with equality by the OR function $\mathrm{OR}(x^n)$. To see this, note that the OR function is monotone and symmetric with a single dominating boundary point 1^n. For this point

$$T_\rho \mathrm{OR}(1^n) = (1 - \delta)^n \cdot 1 + [1 - (1 - \delta)^n] \cdot (-1)$$

which is non-negative if and only if $\delta \in [0, 1 - 2^{-1/n}]$. □

Specific properties of the function, may be used to obtain better sufficient bounds in special cases. For example, suppose that the *sparsity* of $\hat{f}_S$ is s, i.e.,

$$f(x^n) = \sum_{S \in \mathcal{S}} \hat{f}_S \cdot x^S$$

where $\mathcal{S} \subset 2^{[n]}$ and $|\mathcal{S}| = s$. Then, an application of the union bounds leads to

$$\Pr\left[f(X^n) = f(y^n) \mid Y^n = y^n\right] \geq \Pr\left[\bigcap_{S \in \mathcal{S}} X^S = y^S \mid Y^n = y^n\right]$$

$$\geq 1 - \sum_{S \in \mathcal{S}} \Pr\left[X^S \neq y^S \mid Y^n = y^n\right]$$

$$= 1 - \sum_{S \in \mathcal{S}} \frac{1 - \rho^{|S|}}{2}.$$

This probability will be larger than $1/2$ for all $y^n \in \{-1, 1\}^n$ if ρ is larger than the solution to

$$\sum_{S \in \mathcal{S}} \rho^{|S|} = s - 1.$$

Similar conditions can be derived for PTFs (5) of sparsity s.

The extremal property of the OR function noted above may ostensibly be attributed to the fact that it is extremely unbalanced. However, $x_1 \cdot \text{OR}(x_2^n)$ is balanced, and Propositions 3.9 and 4.2 imply that it is ρ-SP for $\rho > 2^{(n-2)/(n-1)} - 1 = 1 - \frac{2\ln(2)}{n} + O(n^{-2})$. The next proposition demonstrates that the statement in Proposition 4.2 holds even if we restrict ourselves to balanced LTFs.

Proposition 4.3. *Any balanced LTF f is ρ-SP for $\rho > 1 - \frac{2\ln(2)}{n} + O(n^{-2})$, and there is no better universal guarantee.*

Proof. Note that the above region is essentially the same as the one in Proposition 4.2, hence one direction is clear. We need to show there exists a balanced function that is not ρ-SP at any point outside this region. To that end, let us introduce the *enlightened dictator* function, defined for $n \geq 3$ to be

$$(9) \qquad \text{E-Dict}(x^n) := \text{sgn}\left((n-2)x_1 + \sum_{i=2}^{n} x_i\right).$$

Evidently, E-Dict(x^n) is determined by the "dictator" x_1, unless all the "subjects" $x_2, \ldots, x_n$ disagree. It is easy to verify that E-Dict(x^n) is a monotone, odd (and hence balanced) function. This function is SP at $y^n = (-1, 1^{n-1})$ if and only if

$$(10) \qquad \Pr(\text{E-Dict}(X^n) = 1 \mid Y^n = y^n) = (1-\delta)^n + \delta(1 - \delta^{n-1}) \geq 1/2.$$

The second derivative of the left-hand side (l.h.s.) above is $n(n-1)((1-\delta)^{n-1} - \delta^{n-2})$, which is non-negative for $\delta \in [0, 1/2]$, hence the l.h.s. is convex inside this interval. It is easy to check that equality in (10) holds for $\delta = \ln(2) \cdot n^{-1} - O(n^{-2})$ and for $\delta = 1/2$, hence by convexity y^n is δ-SP if and only if $\delta < \ln(2) \cdot n^{-1} - O(n^{-2})$, or equivalently, $\rho > 1 - 2\ln(2) \cdot n^{-1} + O(n^{-2})$.[1] $\qquad \square$

4.1. Bounded Degree and Spectral Norm.
Next, we provide a stronger statement that uses the *Fourier-degree* $\mathrm{Deg}(f)$ of the function f, i.e., the maximal degree of the characters appearing in the Fourier representation of f.

Theorem 4.4. *Any function f is ρ-SP for*

$$\rho \geq 1 - \frac{1}{\mathrm{Deg}(f) \cdot \min\left\{\mathrm{Deg}(f), \sum_{S \subseteq [n]} \left|\hat{f}_S\right|\right\}}.$$

Proof. Fix any y^n and think of $T_\rho f(y^n)$ as a polynomial in ρ. Let ρ_0 be the largest root of this polynomial in $[0, 1]$ (if there is one, otherwise $\rho_0 = 0$). Since $T_\rho f(y^n)$ equals $f(y^n) \in \{1, -1\}$ for $\rho = 1$, then by continuity, f is ρ-SP at y^n for any $\rho \geq \rho_0$. By the mean value theorem

$$1 = T_1 f(y^n) - T_{\rho_0} f(y^n) = (1 - \rho_0) \left.\frac{\mathrm{d}}{\mathrm{d}\rho} T_\rho f(y^n)\right|_{\rho=\tilde{\rho}}$$

for some $\tilde{\rho} \in [\rho_0, 1]$, and so

$$(11) \qquad \rho_0 \leq 1 - \frac{1}{\max_{\rho \in [0,1]} \left|\frac{\mathrm{d}}{\mathrm{d}\rho} T_\rho f(y^n)\right|},$$

and so a bound on ρ_0 may be obtained by bounding the derivative. To that end, recall that Markov brothers' inequality [GM99, Theorem 1.1] states that for any real polynomial $P(t)$ of degree k

$$\max_{t \in [-1,1]} \left|\frac{\mathrm{d}}{\mathrm{d}t} P(t)\right| \leq k^2 \cdot \max_{t \in [-1,1]} |P(t)|,$$

and that Bernstein's inequality [GM99, Theorem 1.2] states that for any complex polynomial $Q(z)$ of degree k,

$$\max_{|z| \leq 1} \left|\frac{\mathrm{d}Q(z)}{\mathrm{d}z}\right| \leq k \cdot \max_{|z| \leq 1} |Q(z)|.$$

The claim then follows from (11) by noting that the degree of $T_\rho f$ as a polynomial in ρ equals the Fourier degree $\mathrm{Deg}(f)$, and the bound

$$|T_\rho f(y^n)| = \left|\sum_{S \subseteq [n]} \rho^{|S|} \cdot \hat{f}_S \cdot x^S\right|$$

---

[1]Using Proposition 4.1 it can be verified that this the true range for which E-Dict($\cdot$) is SP.

$$\leq \sum_{S \subseteq [n]} |\hat{f}_S|$$

for any $\rho \in (0, 1]$. □

Theorem 4.4 significantly improves on Theorem 4.2 whenever $\mathrm{Deg}(f) \ll \sqrt{n}$, e.g., for n-dimensional functions f that can be computed by a decision tree of depth $k \ll n$, in which case $\mathrm{Deg}(f) \leq k$ [O'D14, Proposition 3.16]. Functions with low *spectral norm* $\sum_{S \subseteq [n]} |\hat{f}_S|$ are discussed in [STlV17] and references therein.

4.2. **Friendly Neighbors.** Given a function f, we say that a point x^n has a *radius-d friendly neighborhood* w.r.t. f if there exists some y^n of distance at most d that agrees with x^n, namely, where $d_H(x^n, y^n) \leq d$ and $f(x^n) = f(y^n)$.

Proposition 4.5. *Suppose f is ρ-SP for all $\rho > 1 - \varepsilon$, and $n > \max\{2\varepsilon^{-1}, \gamma\}$ where γ is a universal constant. Then each point in $\{-1, 1\}^n$ has a radius-2 friendly neighborhood w.r.t. f.*

Proof. Suppose toward contradiction that all the neighbors at Hamming distance 1 and 2 from some y^n disagree with it. This implies that

$$\Pr\left(f(X^n) \neq f(Y^n) \mid Y^n = y^n\right) \geq \binom{n}{1} \delta (1-\delta)^{n-1} + \binom{n}{2} \delta^2 (1-\delta)^{n-2}$$
$$= (1-\delta)^{n-2} n \delta \left((1-\delta) + \frac{(n-1)}{2} \delta \right).$$

Choosing $\delta = \frac{\alpha}{n}$, and assuming that $n > \frac{2\alpha}{\varepsilon}$ so that we are in the SP region, yields

$$\Pr\left(f(X^n) \neq f(Y^n) | Y^n = y^n\right) \geq \left(1 - \frac{\alpha}{n}\right)^{n-2} \alpha \left(1 + \frac{\alpha}{2} - \frac{3\alpha}{2n}\right)$$
$$\geq \left(1 - \frac{\alpha}{n}\right)^{n-2} \cdot \left(\alpha + \frac{\alpha^2}{2}\right) - O\left(\frac{1}{n}\right)$$
$$= e^{-\alpha} \cdot \left(\alpha + \frac{\alpha^2}{2}\right) - O\left(\frac{1}{n}\right).$$

One can check that, e.g., for $\alpha = 1$, $(\alpha + \frac{\alpha^2}{2})e^{-\alpha} > 1/2$, and so f cannot be SP if n is larger than some universal constant, in contradiction. □

Hence, for a function to be SP even slightly below the guaranteed high correlation threshold of $\rho > 1 - \frac{2\ln(2)}{n} + O(n^{-2})$, every point must admit a radius-2 friendly neighborhood. The OR function, e.g., does not satisfy this property. Furthermore, this result is tight: for the largest character $x^{[n]} = \prod_{i=1}^{n} x_i$, which is USP, the distance-1 neighbors of each point do not agree with it.

The following corollary, which is not directly related to self-predictability, is obtained by combining Theorem 4.4 and Proposition 4.5.

Corollary 4.6. *If* $\mathrm{Deg}\, f < \sqrt{n/2}$ *and* n *is larger than a universal constant, then each point in* $\{-1,1\}^n$ *has a radius-2 friendly neighborhood w.r.t.* f.

5. Low Correlation Self Predicting (LCSP) Functions

In this section we discuss LCSP functions, i.e., functions that are ρ-SP for any $\rho < \rho^*$ for some $\rho^* > 0$. Note that any USP function is trivially also LCSP, hence all our LCSP necessary conditions will apply to USP functions verbatim.

5.1. **LCSP and Spectral Threshold Functions.** Let the *minimal level* of a function f be defined as

$$\mathrm{Lev}(f) := \min \left\{ k \in [n] : W^k[f] > 0 \right\},$$

and let

$$f_{\mathrm{Lev}}(x^n) := \sum_{S:|S|=\mathrm{Lev}(f)} \hat{f}_S x^S.$$

We say that f is *weakly spectral threshold (WST)* if $f_{\mathrm{Lev}}(x^n) \cdot f(x^n) \geq 0$ for all x^n, i.e., the sign of both functions agree whenever $f_{\mathrm{Lev}} \neq 0$. We say that f is *strongly spectral threshold (SST)* if it is WST and f_{Lev} is never zero.

For an LTF $f = \mathrm{sgn}(a_0 + \sum_{i=1}^n a_i x_i)$, the Fourier coefficients $(\hat{f}_\phi, \hat{f}_1, \ldots, \hat{f}_n)$ are called *Chow parameters*, and, as is well-known [Cho61, Tan61], these parameters unambiguously determine the LTF. The *Chow-parameters problem* [OS11] is to find coefficients a_0^n defining the LTF given the Chow parameters. It can be seen that in case of balanced LTFs, SST functions are exactly the LTFs for which a solution to the Chow-parameters problem is exactly the Chow parameters themselves.

Proposition 5.1. *SST implies LCSP. Conversely, LCSP implies WST.*

Proof. The optimal predictor for f satisfies

$$\mathrm{sgn}\, T_\rho f(x^n) = \mathrm{sgn} \left(\rho^{\mathrm{Lev}(f)} \cdot \sum_{s:|S|\geq\mathrm{Lev}(f)} \rho^{|S|-\mathrm{Lev}(f)} \hat{f}_S x^S \right)$$

$$= \mathrm{sgn}\left(f_{\mathrm{Lev}}(x^n) + O(\rho) \right).$$

Thus, $\mathrm{sgn}\, T_\rho f(x^n) = \mathrm{sgn}\, f_{\mathrm{Lev}}(x^n)$ for any ρ small enough whenever $f_{\mathrm{Lev}}(x^n) \neq 0$. If f is SST $f_{\mathrm{Lev}}(x^n)$ never vanishes, and hence $f(x^n) = \mathrm{sgn}\, f_{\mathrm{Lev}}(x^n) = \mathrm{sgn}\, T_\rho f(x^n)$, implying LCSP. Conversely, if f is LCSP, then $f(x^n) = \mathrm{sgn}\, T_\rho f(x^n) = \mathrm{sgn}\, f_{\mathrm{Lev}}(x^n)$ unless f_{Lev} vanishes, implying WST. $\square$

An immediate consequence of Proposition 5.1 is:

Corollary 5.2. *An LCSP function is either balanced or constant.*

Proof. Suppose f is LCSP and unbalanced. Then $\text{Lev}[f] = 0$ and $\hat{f}_\phi \neq 0$, and by Proposition 5.1 it must be WST. Hence $f = \text{sgn}\,\hat{f}_\phi \in \{-1, 1\}$ must be constant. $\square$

It is also interesting to note the following dichotomy:

Corollary 5.3. *Let f be an LCSP function. Then either $W^1[f] = 0$ or $W^1[f] \geq 1/2$.*

Proof. If $0 < W^1[f] < 1/2$ then Proposition 3.17 implies that

$$\frac{\text{Stab}_\rho^*[f]}{\text{Stab}_\rho[f]} > 1$$

for all sufficiently small ρ, and so f cannot be LCSP. $\square$

This result resembles the claim that $W^1[f] \geq 1/2$ for LTFs [O'D14, Theorem 5.2]. Note however that the above claim holds for LCSP functions that are not LTFs but do have energy on the first level. Next, recall that Proposition 3.13 states that a function is ρ-SP if and only if $\|T_\rho f\|_1 = \text{Stab}_\rho^*[f] = \text{Stab}_\rho[f] = \|T_{\sqrt{\rho}} f\|_2^2$. A similar property holds for f_{Lev} if the function is LCSP.

Corollary 5.4. *If f is LCSP then $\|f_{\text{Lev}}\|_1 = \|f_{\text{Lev}}\|_2^2$.*

Proof. f must be WST by Proposition 5.1, and so Plancherel's identity implies that

$$\mathbb{E}\,|f_{\text{Lev}}(X^n)| = \mathbb{E}\,(f_{\text{Lev}}(X^n) \cdot f(X^n))$$
$$= \langle f_{\text{Lev}}, f \rangle$$
$$= \sum_{S:|S|=\text{Lev}[f]} \hat{f}_S^2$$
$$= \mathbb{E}\,(f_{\text{Lev}}^2(X^n)).$$

$\square$

The following two examples show that the distinction between WST and SST in Proposition 5.1 is necessary.

Example 5.5 (LCSP does not imply SST)**.** Consider the balanced LTF with $n = 4$ and coefficients $a_1^4 = (2, 1, 1, 1)$. This is a Majority function with a tie-breaking input. It can be verified by direct computation that this function is USP, hence also LCSP. However, its level-1 Fourier coefficients are $(\frac{3}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. Hence, while it is clearly WST, it is not SST as there are 2 inputs for which $f_{\text{Lev}}(x^n) = 0$.

Example 5.6 (WST does not imply LCSP)**.** The balanced LTF with $n = 9$ and coefficients $a_1^9 = (1, 5, 16, 19, 25, 58, 68, 91, 94)$ can be verified to be WST, but not LCSP. It is ρ-SP only for $\rho > 0.577$. This example was found by analyzing LTFs with randomly drawn coefficients.

The following example shows that the SST property is limited to the low-correlation regime only.

Example 5.7 (SST does not imply USP). The LTF of Example 3.3 is SST, but as was shown there, is not USP. Thus, while an SST is always LCSP, it is not necessarily USP.

We note in passing that there are SST and WST functions outside Majority that are USP.

Example 5.8. The LTF in Example 3.7 is SST and USP, while the balanced LTF with $n = 9$ and coefficients $a_1^9 = (1, 1, 1, 3, 3, 3, 5, 5, 7)$ is WST and USP ($f_{\text{Lev}} = 0$ for 30 inputs), but not SST.

Next, using Proposition 5.1, we can show that the largest coefficients of an LCSP LTF cannot be too distinct.

Proposition 5.9. *Let f be an LTF that depends on all its n variables. Let a and b be its first and second largest coefficients in absolute values, respectively, in some representation of f. If f is LCSP then $\left|\frac{a}{b}\right| < \sqrt{2n \ln(2n)} + 1$.*

Proof. Assume without loss of generality that $a_1 \geq a_2 \geq \cdots \geq a_n > 0$. Recall also that by Corollary 5.2 we know that $a_0 = 0$. Since f is monotone, its level-1 Fourier coefficients equal influences [O'D14, Proposition 2.21], i.e.,

$$\hat{f}_k = \text{Inf}_k[f] \tag{12}$$

$$:= \Pr\left(f(X^n) \neq f(X_1^{k-1}, -X_k, X_{k+1}^n)\right)$$

$$= \Pr\left(\left|\sum_{i \neq k} a_i X_i\right| < a_k\right). \tag{13}$$

Assume without loss of generality that $a_2 = 1$, and write $a := a_1$. For brevity, also write $Z := \sum_{i=3}^n a_i X_i$ and $X := X_1$. Then, from the symmetry of Z,

$$\hat{f}_1 = \Pr(|X + Z| \leq a)$$

$$= \Pr(|1 + Z| \leq a)$$

$$\geq \Pr(|Z| < a - 1),$$

and

$$\hat{f}_2 = \Pr(|aX + Z| \leq 1)$$

$$\leq \Pr(a - 1 \leq |Z| \leq a + 1)$$

$$\leq \Pr(|Z| \geq a - 1).$$

Hence,

$$\frac{\hat{f}_1}{\hat{f}_2} \geq \frac{1 - \Pr(|Z| \geq a - 1)}{\Pr(|Z| \geq a - 1)}.$$

Since $|a_i| \leq 1$ for $3 \leq i \leq n$, and assuming toward contradiction that $a > \sqrt{(2n - 2) \ln 2n} + 1$, Hoeffding's inequality implies that

$$\Pr(|Z| \geq a - 1) < 1/n,$$

and so $\hat{f}_1/\hat{f}_2 > n - 1$. Noting that $a_i \geq a_j$ implies $\hat{f}_i \geq \hat{f}_j$, we also have that $\hat{f}_1/\hat{f}_i \geq n - 1 + \varepsilon$ for any $i > 1$, for $\varepsilon > 0$ small enough. By Proposition 5.1, f is WST, i.e., $f(x^n) = \operatorname{sgn}\sum_{i=1}^n \hat{f}_i x_i$ whenever the right-hand side (r.h.s.) is nonzero. This representation and the bounds on the ratios $\hat{f}_1/\hat{f}_i$ from above imply that $f(x^n) = x_1$ must hold. This, however, contradicts the assumption that f depends on all the variables. $\qquad\square$

For example, the enlightened dictator function E-Dict$(\cdot)$ (9) has first-to-second co-efficient ratio of $n - 2$, and thus cannot be LCSP. It should be noted however, that E-Dict$(\cdot)$ can also be written as an LTF with coefficients E-Dict$(\cdot) = (\sqrt{n}, 1, c, c, \ldots, c)$ where $c = \frac{\sqrt{n} - 1 + \varepsilon}{n - 2}$ for some $\varepsilon > 0$. When given in this form, Proposition 5.9 is incapable of ruling it out from being SP. Nonetheless, it is easy to verify that LTFs of coefficients $(c, 1, 1, \ldots, 1)$ for $c < n - 2$ and $c = \Omega(n)$, must have $a_2 = a_3 \cdots = a_n$ in any valid representation, and thus the first-to-second-coefficient ratio is always $\Omega(n)$.

5.2. **LTF Approximation.** The WST condition can be leveraged to show that a LCSP function can typically be well approximated by an LTF. Specifically:

Theorem 5.10. *An LCSP f is $\sqrt{\frac{2}{\pi n_f}}$-close to an LTF, where $n_f := |\{i \in [n] : \hat{f}_i \neq 0\}|$.*

Corollary 5.11. *A monotone LCSP function that depends on all its coordinates is $\sqrt{\frac{2}{\pi n}}$-close to an LTF.*

To prove Theorem 5.10 we first establish the following technical lemma. We state it in a slightly more general form than we actually need.

Lemma 5.12. *Let $a^n \in \mathbb{R}^n$ be a vector of nonzero coefficients. Then for any $b \in \mathbb{R}$*

$$\Pr\left(\left|\sum_{i=1}^n a_i X_i - b\right| < \min_{k \in [n]} |a_k|\right) \leq 2^{-n}\binom{n}{\lfloor n/2 \rfloor} \leq \sqrt{\frac{2}{\pi n}}.$$

Proof. Write $a = \min|a_k|$ and let

$$\mathcal{A} := \left\{x^n \in \{-1, 1\}^n : \left|\sum_{i=1}^n a_i x_i - b\right| < a\right\}.$$

It is easy to see that $\mathcal{A}$ forms an antichain w.r.t. the partial order $\preceq$ on $\{-1, 1\}^n$, i.e., that there are no two distinct $x^n, y^n \in \mathcal{A}$ such that $x^n \preceq y^n$. This holds simply since for such a pair it must hold that

$$\left| \sum_{i=1}^{n} a_i y_i - \sum_{i=1}^{n} a_i x_i \right| \geq 2a.$$

An antichain w.r.t. $\preceq$ is called a *Sperner family*, and *Sperner's theorem* [AS04, Maximal Antichains, Corollary 2] shows that

$$|\mathcal{A}| \leq \binom{n}{\lfloor n/2 \rfloor}$$

concluding the proof. $\qquad\square$

Proof of Theorem 5.10. Assume $\text{Lev}[f] = 1$ (trivial otherwise), and define $g(x^n) = \text{sgn}(\sum_{i=1}^{n} \hat{f}_i x_i)$. Let $\mathcal{A} := \{x^n \in \{-1, 1\}^n : g(x^n) = 0\}$. Using Lemma 5.12, we have that

$$\Pr(X^n \in \mathcal{A}) \leq \sqrt{\frac{2}{\pi n_f}}.$$

Since f is LCSP then by Proposition 5.1 is it also WST, and hence $f(x^n) = g(x^n)$ for all $x^n \notin \mathcal{A}$. By slightly perturbing the coefficients of g, one can clearly obtain a "legal" LTF $\tilde{g}$ that takes values only in $\{-1, 1\}$ and still agrees with f for all $x^n \notin \mathcal{A}$. The distance between f and $\tilde{g}$ is therefore at most $|\mathcal{A}|/2^n$. $\qquad\square$

5.3. **Chow Distance.** The *Chow distance* between two Boolean functions f and g is defined as

$$d_{\text{Chow}}(f, g) := \left(\sum_{i \in [n]} \left(\hat{f}_i - \hat{g}_i \right)^2 \right)^{1/2}.$$

It was shown in [OS11, Prop. 1.5, Th. 1.6] that for any f and g

$$\frac{1}{4} d^2_{\text{Chow}}(f, g) \leq \text{Dist}(f, g) \leq \tilde{O}\left(\frac{1}{\sqrt{-\log d_{\text{Chow}}(f, g)}} \right),$$

where for $q < 1$, $\tilde{O}(q)$ means $O(q \cdot \log^c(1/q))$ for some absolute constant c.

For LCSP LTF functions, the upper bound can be generally improved. We will state our result for the case where one of the functions is SST, though it can be somewhat cumbersomely extended to the case where none of them is. Let $\text{Gap}[f]$ be the minimal positive value of $\sum_{i=1}^{n} \hat{f}_i x_i$ over the Hamming cube (with $\text{Gap}[f] = 0$ if all the $\hat{f}_i$'s are zero).

Theorem 5.13. *Let f and g be two balanced LCSP functions that depend on all n variables, and assume that f is SST. Then*

$$\text{Dist}(f, g) \leq \frac{d^2_{\text{Chow}}(f, g)}{2 \, \text{Gap}[f]}.$$

Proof of Theorem 5.13. Let

$$\mathcal{B} := \{x^n \in \{-1, 1\}^n : f(x^n) \neq g(x^n)\} \, .$$

Then,

$$(14) \qquad d^2_{\text{Chow}}(f, g) = \mathbb{E} \left((f(X^n) - g(X^n)) \cdot \sum_{i \in [n]} \left(\hat{f}_i - \hat{g}_i \right) X_i \right)$$

$$(15) \qquad = 2\mathbb{E} \left(\left| \sum_{i \in [n]} \left(\hat{f}_i - \hat{g}_i \right) X_i \right| \cdot \mathbb{1} (X^n \in \mathcal{B}) \right)$$

$$(16) \qquad \geq 2 \, \text{Gap}[f] \cdot \Pr \left(X^n \in \mathcal{B} \right),$$

where (14) follows from linearity of expectation and the definition of the Fourier coefficients, (15) holds since both f and g are WST by virtue of Proposition 5.1 and so for all $x^n \in \mathcal{B}$, $|f(X^n) - g(X^n)| = 2$ and $\text{sgn}[\sum_{i \in [n]} (\hat{f}_i - \hat{g}_i) X_i] = \text{sgn}[f(X^n) - g(X^n)]$. Finally, (16) holds by noting that f is SST and g is WST. Thus, whenever $f(x^n) > g(x^n)$ then $\sum_{i \in [n]} \hat{f}_i X_i > \text{Gap}[f]$ and $\sum_{i \in [n]} \hat{g}_i X_i \leq 0$ (and similarly for $f(x^n) < g(x^n)$). $\qquad \square$

Equations (12)-(13) and Lemma 5.12 imply that $\text{Gap}[\text{Maj}] \leq \sqrt{\frac{2}{\pi n}}$. Since Majority is SST, we have:

Corollary 5.14. *For odd n and any LCSP function g,*

$$\frac{1}{4} \cdot d^2_{\text{Chow}}(\text{Maj}, g) \leq \text{Dist}(\text{Maj}, g) \leq \sqrt{\frac{\pi n}{8}} \cdot d^2_{\text{Chow}}(\text{Maj}, g).$$

6. STABILITY-BASED CONDITIONS

In this section we provide simple necessary conditions for a function to be ρ-SP, in terms of its stability and Fourier coefficients.

Proposition 6.1. *If f is ρ-SP then*

$$\text{Stab}_\rho[f] \geq \max_{S \subseteq [n]} \rho^{|S|} |\hat{f}_S|.$$

Proof. If f is ρ-SP, then $\text{Stab}_\rho[f] = \text{Stab}^*_\rho[f]$. Letting $T \subseteq [n]$, the strong stability can be lower bounded as follows:

$$\text{Stab}^*_\rho[f] = \mathbb{E} \left| \sum_{S \subseteq [n]} \rho^{|S|} \cdot \hat{f}_S \cdot Y^S \right|$$

$$= \mathbb{E}\left(\left| \sum_{S \subseteq [n]} \rho^{|S|} \cdot \hat{f}_S \cdot Y^S \right| \cdot |Y^T| \right)$$

$$= \mathbb{E}\left(\left| \sum_{S \subseteq [n]} \rho^{|S|} \cdot \hat{f}_S \cdot Y^S \cdot Y^T \right| \right)$$

$$\geq \left| \mathbb{E}\left(\sum_{S \subseteq [n]} \rho^{|S|} \cdot \hat{f}_S \cdot Y^S \cdot Y^T \right) \right|$$

$$= |\rho^{|T|} \cdot \hat{f}_T|.$$

The proof is completed by optimizing over T. $\qquad \square$

Example 6.2. When f is the OR function, we have

$$\max_{S \subseteq [n]} \rho^{|S|} |\hat{f}_S| = |\hat{f}_\phi| = 1 - 2^{1-n}.$$

It is easy to verify that

$$\Pr\left(f(X^n) = f(Y^n) \right) = 1 - 2^{1-n} \cdot \left(1 - (1 - \delta)^n \right),$$

and using $\rho = 1 - 2\delta$

$$\mathrm{Stab}_\rho[f] = 2 \cdot \Pr\left(f(X^n) = f(Y^n) \right) - 1$$

$$= 1 - 2^{2-n} \cdot \left(1 - \left(\frac{1+\rho}{2} \right)^n \right).$$

Then, OR is ρ-SP only when $\mathrm{Stab}_\rho[\mathrm{OR}] \geq 1 - 2^{1-n}$, which can be seen to be equivalent to $\rho \geq 2^{(n-1)/n} - 1$. This is the same result that can be obtained by direct computation (see Proposition 4.2), and so the bound of Proposition 6.1 is tight in this case. Furthermore, we may deduce again the result of Corollary 5.2:

Corollary 6.3. *An LCSP function is either balanced or constant.*

Proof. If f is ρ-SP then

$$\mathrm{Stab}_\rho[f] = \sum_{S \subseteq [n]} \rho^{|S|} |\hat{f}_S|^2 \geq |\hat{f}_\phi|.$$

As $\rho \downarrow 0$, this bound implies that $|\hat{f}_\phi|^2 \geq |\hat{f}_\phi|$, and as $|\hat{f}_\phi| \leq 1$, this is only possible when either $\hat{f}_\phi = 0$ or $|\hat{f}_\phi| = 1$. $\qquad \square$

More generally, we have the following:

Corollary 6.4. *If f is LCSP then*

$$W^{\mathrm{Lev}[f]}[f] \geq \max_{S \subseteq [n]:\, |S| = \mathrm{Lev}(f)} |\hat{f}_S|.$$

Specifically, if f is also monotone, this bound reads

$$W^1[f] \geq \max_{i \in [n]} \hat{f}_i = \max_{i \in [n]} \mathrm{Inf}_i[f],$$

where the r.h.s. is the so-called maximal influence *of f.*

When $\mathrm{Deg}(f) < n$, another bound of the form of Proposition 6.1 can be derived using the following implication of hypercontractivity [Bon70, Gro75]: When $f : \{-1, 1\}^n \to \mathbb{R}$ has $\mathrm{Deg}(f) = k$ then $\|f\|_2 \leq e^k \cdot \|f\|_1$ [O'D14, Theorem 9.22].

Proposition 6.5. *If f is ρ-SP and $\mathrm{Deg}(f) = k$ then*

$$\mathrm{Stab}_\rho[f] \geq e^{-k} \cdot \sqrt{\mathrm{Stab}_{\rho^2}[f]}.$$

Proof. As in the proof of Proposition 6.1, we lower bound

$$\mathbb{E}\left| \sum_{S \subseteq [n]} \rho^{|S|} \cdot \hat{f}_S \cdot Y^S \right| = \|T_\rho f\|_1$$

$$(17) \qquad\qquad\qquad \geq e^{-k} \cdot \|T_\rho f\|_2$$

$$= e^{-k} \cdot \sqrt{\langle T_\rho f, T_\rho f \rangle}$$

$$(18) \qquad\qquad\qquad = e^{-k} \cdot \sqrt{\langle T_{\rho^2} f, f \rangle}$$

$$= e^{-k} \cdot \sqrt{\mathrm{Stab}_{\rho^2}[f]}$$

where (17) is since $\mathrm{Deg}(f) = \mathrm{Deg}(T_\rho f) = k$, and (18) is since $T_\rho f$ is a self-adjoint operator. $\square$

The last proof implies for a degree k, ρ-SP function f

$$e^{-k} \cdot \sqrt{\mathrm{Stab}_{\rho^2}[f]} \leq \mathrm{Stab}_\rho[f] \leq \sqrt{\mathrm{Stab}_{\rho^2}[f]}.$$

It can be observed that even for a given degree k, neither of the bounds in Propositions 6.1 and 6.5 subsumes the other.

7. SHARP THRESHOLD AT HIGH CORRELATION

As we have seen, all functions are ρ-SP when $\rho > 1 - \frac{2 \ln 2}{n} + O(n^{-2})$. In this section, we show that when the correlation is reduced ever so slightly to $\rho \approx 1 - \frac{2}{n}$, the fraction of SP functions becomes double-exponentially small.

Theorem 7.1. *For any $\alpha > 1$, the fraction of ρ-SP functions for $\rho = 1 - \frac{2\alpha}{n}$ is at most $\exp(-2^{n \cdot E(\alpha) + o(n)})$, where*

$$E(\alpha) := \min \left\{ \frac{1}{2}, h\left(\frac{\alpha - 1}{2\alpha} \right) \right\}$$

and $h(t) := -t \log(t) - (1 - t) \cdot \log(1 - t)$ is the binary entropy function.

The fact that ρ-SP functions are rare is not limited to the $\rho = 1 - O(\frac{1}{n})$ regime, yet a different technique is needed in order to establish this in other regimes. We next demonstrate how a similar phenomenon holds in a high correlation regime where ρ is fixed. Let η_δ be the minimal $\eta > 0$ such that

$$\frac{1}{2} \log \frac{1}{\delta^2 + (1 - \delta)^2} < \min \left\{ \log \frac{1}{1 - \delta}, d(\eta||\delta) \right\}$$

holds, where $d(p||q) := p \log \frac{p}{q} + (1 - p) \log \frac{1-p}{1-q}$ is the binary divergence function. It can be verified that $\eta_\delta < 1/4$ for any $\delta < \delta_{\max} \approx 0.0974$.

Theorem 7.2. *For any $\delta \in (0, \delta_{\max})$, the fraction of ρ-SP functions for $\rho = 1 - 2\delta$ is at most $\exp\left(-2^{n[1-h(2\eta_\delta)]-o(n)}\right)$.*

We begin with the proof of Theorem 7.1.

Proof of Theorem 7.1. In this proof we find it more convenient to work with a δ and $\{0, 1\}$ convention. We begin by deriving a sufficient condition for a function to be non-ρ-SP at any fixed y^n. This condition depends only on local values of the function, up to a Hamming distance of $\log n$ from y^n, and is tailored to the regime of $\delta = \Theta(1/n)$. Specifically, we show that for a random choice of function, the probability that our condition is satisfied decays exponentially with n, and we derive an upper bound on the associated exponent. Then, since the resulting exponent is smaller than 1, we conclude that the expected number of non-ρ-SP points for a random function is exponentially large. This fact in itself, however, is not sufficient since there are statistical dependencies between different points in the Hamming cube. Nonetheless, Janson's theorem [AS04, Theorem 8.1.1] along with the aforementioned "locality" of the sufficient condition allow us to prove that the probability that all points in the Hamming cube are ρ-SP is only double-exponentially small.

We proceed to prove the local condition for non-ρ-SP-ness. To that end, let us denote the *shell of radius d around $x^n \in \{0,1\}^n$* by

$$\mathcal{S}(x^n, d) := \{\tilde{x}^n : d_H(x^n, \tilde{x}^n) = d\}.$$

For any function $f : \{0, 1\}^n \to \{0, 1\}$, let the *$d$-shell bias of f* be

$$\beta_{d,f}(x^n) := \frac{1}{|\mathcal{S}(x^n, d)|} \sum_{\tilde{x}^n \in \mathcal{S}(x^n, d)} f(\tilde{x}^n).$$

Fix $\eta > 0$ and some y^n. Without loss of generality, we assume below that $f(y^n) = 0$. Define the set of functions

$$\mathcal{B}_\eta(y^n, 1) = \{f : \beta_{1,f}(y^n) \geq 1 - \eta\},$$

and for $2 \leq d \leq \ell$, the sets

$$\mathcal{B}_\eta(y^n, d) = \{f : \beta_{d,f}(y^n) \geq 1/2\},$$

where $\ell \geq 3$. We say that y^n is *bad* for f if $f \in \mathcal{B}_\eta(y^n)$, where

$$\mathcal{B}_\eta(y^n) := \bigcap_{d=1}^{\ell} \mathcal{B}_\eta(y^n, d).$$

Now, for any $n > \ell$, setting $\delta = \frac{\alpha}{n}$, any $f \in \mathcal{B}_\eta(y^n)$ satisfies:

$$\Pr\left(f(X^n) \neq f(Y^n) \mid Y^n = y^n\right)$$

$$\geq \sum_{d=1}^{\ell} \beta_{d,f}(y^n) \binom{n}{d} \delta^d (1-\delta)^{n-d}$$

$$\geq (1-\delta)^n \cdot \sum_{d=1}^{\ell} \beta_{d,f}(y^n) \binom{n}{d} \delta^d$$

$$\geq \left(1 - \frac{\alpha}{n}\right)^n \cdot \left(1 - \frac{\ell}{n}\right)^\ell \cdot \left(\sum_{d=1}^{\ell} \frac{\beta_{d,f}(y^n)}{d!} \cdot \alpha^d\right)$$

$$\geq \left(1 - \frac{\alpha}{n}\right)^n \cdot \left(1 - \frac{\ell}{n}\right)^\ell \cdot \left((1-\eta) \cdot \alpha + \frac{1}{2} \sum_{d=2}^{\ell} \frac{\alpha^d}{d!}\right)$$

$$(19) \qquad = \left(1 - \frac{\alpha}{n}\right)^n \cdot \left(1 - \frac{\ell}{n}\right)^\ell \cdot \left((1-\eta) \cdot \alpha + \frac{1}{2}\left(e^\alpha - 1 - \alpha - \sum_{d=\ell+1}^{\infty} \frac{\alpha^d}{d!}\right)\right).$$

Taking ℓ to be $\Omega(1)$ and $o(n)$, say $\ell = \log n$, (19) tends to

$$\frac{1}{2} + \left(\frac{1}{2} - \eta\right) \alpha e^{-\alpha} - \frac{1}{2} e^{-\alpha}$$

as $n \to \infty$. Let

$$\eta_\alpha := \frac{\alpha - 1}{2\alpha}.$$

Clearly, η_α is monotonically increasing for $\alpha > 0$, where $\lim_{\alpha \downarrow 1} \eta_\alpha = 0$, and $\lim_{\alpha \uparrow \infty} \eta_\alpha = 1/2$. Setting $\eta \in (0, \eta_\alpha)$ guarantees that (19) is larger than $1/2$ for all large enough n. Hence, for such a choice,

$$\Pr\left(f(X^n) \neq f(Y^n) \mid Y = y^n\right) > 1/2,$$

and so

$$\{f \in \mathcal{B}_\eta(y^n)\} \subseteq \{f \text{ is not } \rho\text{-SP at } y^n\}.$$

Let us now choose f uniformly at random over all Boolean functions on $\{0,1\}^n$, and lower bound the probability that f is ρ-SP at y^n. To that end, note that Chernoff's

bound implies that

$$\text{(20)} \qquad \Pr\left(\beta_{1,f}(y^n) \geq 1 - \eta\right) = 2^{-n(1-h(\eta))+o(n)},$$

and symmetry implies that

$$\Pr\left(\beta_{d,f}(y^n) \geq 1/2\right) \geq 1/2,$$

for $2 \leq d \leq \ell = \log n$. By independence,

$$
\begin{aligned}
\Pr\left(f \in \mathcal{B}_\eta(y^n)\right) &= \prod_{d=1}^{\log n} \Pr\left(f \in \mathcal{B}_\eta(y^n, d)\right) \\
&= 2^{-n(1-h(\eta))+o(n)} \cdot 2^{-(\log n - 1)} \\
&= 2^{-n(1-h(\eta))+o(n)},
\end{aligned}
$$

and so

$$\Pr\left(f \text{ is not } \rho\text{-SP at } y^n\right) \geq 2^{-n(1-h(\eta))+o(n)}.$$

This completes the proof of the local bound.

We now proceed to the global behavior of the number of non-ρ-SP points. Let us first upper bound the probability $\Pr(f \in \mathcal{E})$ where

$$\mathcal{E} := \bigcap_{y^n \in \{0,1\}^n} \mathcal{B}_\eta^c(y^n).$$

This in turn will serve as an upper bound for the probability that the function we draw is ρ-SP for the aforementioned ρ. To that end, note that if f is ρ-SP for $\rho = 1 - 2\alpha \cdot n^{-1}$ then it must be that f has no bad inputs, i.e., $f \in \mathcal{E}$. Furthermore, note that the expected number of "bad" inputs is given by

$$\mu := 2^n \cdot \Pr\left(f \in \mathcal{B}_\eta(y^n)\right) = 2^{nh(\eta)+o(n)}.$$

If the number of bad inputs had been Poisson distributed with mean μ, then

$$\Pr\left(f \in \mathcal{E}\right) = e^{-\mu} = \exp\left(-2^{n \cdot h(\eta)+o(n)}\right).$$

However, the events $\mathcal{B}_\eta(x^n)$ and $\mathcal{B}_\eta(y^n)$ are dependent whenever $d_H(x^n, y^n) \leq 2\ell$. Nonetheless, Janson's correction [AS04, Theorem 8.1.1] implies that

$$\Pr\left(f \in \mathcal{E}\right) \leq e^{-\mu + \frac{\Delta}{2}},$$

where Δ is a correction term that depends on joint probability of dependent bad events $\Pr(f \in \mathcal{B}_\eta(x^n) \cap \mathcal{B}_\eta(y^n))$. We next show that $\Delta \to 0$ as $n \to \infty$ exponentially fast, as long as $\eta \in (0, h^{-1}(1/2))$. Once this is established, one can set $\eta = \min\{\eta_\alpha, h^{-1}(1/2)\}$ to obtain

$$\Pr\left(f \in \mathcal{E}\right) \leq \exp\left(-2^{n \cdot h(\eta)+o(n)}\right),$$

and the theorem follows.

To complete the proof, it remains to show that $\Delta \to 0$ exponentially fast. Let us denote $x^n \sim y^n$ whenever the events $\{f \in \mathcal{B}_\eta(x^n)\}$ and $\{f \in \mathcal{B}_\eta(y^n)\}$ are statistically dependent. The term required for Janson's theorem is then given by

$$(21) \qquad \Delta := \sum_{x^n \sim y^n} \Pr\left(f \in \mathcal{B}_\eta(x^n) \cap \mathcal{B}_\eta(y^n)\right).$$

Let us analyze the probability in (21) under the assumption that $f(x^n) = f(y^n) = 0$. It will be evident that all other three cases for $(f(x^n), f(y^n))$ can be analyzed in the same way and lead to essentially the same result. Bayes rule implies that

$$\Pr\left(f \in \mathcal{B}_\eta(x^n) \cap \mathcal{B}_\eta(y^n) \mid f(x^n) = f(y^n) = 0\right)$$
$$= \Pr\left(f \in \mathcal{B}_\eta(x^n, 1) \mid f(x^n) = f(y^n) = 0\right)$$
$$\times \Pr\left(f \in \mathcal{B}_\eta(y^n, 1) \mid f(x^n) = f(y^n) = 0, f \in \mathcal{B}_\eta(x^n, 1)\right)$$
$$\times \Pr\left(\bigcap_{d=2}^{\ell} \{\{f \in \mathcal{B}_\eta(x^n, d)\} \cap \{f \in \mathcal{B}_\eta(y^n, d)\}\} \mid \right.$$
$$(22) \qquad \left. f(x^n) = f(y^n) = 0, f \in \mathcal{B}_\eta(x^n, 1) \cap \mathcal{B}_\eta(y^n, 1)\right).$$

For the first probability on the r.h.s. of (22), we note that if $d_H(x^n, y^n) \geq 2$ then $\mathcal{S}(x^n, 1) \cap \{x^n, y^n\} = \phi$ and (20) holds. Otherwise, if $d_H(x^n, y^n) = 1$ then $\mathcal{S}(x^n, 1) \cap \{x^n, y^n\} = y^n$. In that case,

$$\Pr\left(f \in \mathcal{B}_\eta(x^n, 1) \mid f(x^n) = f(y^n) = 0\right)$$
$$= \Pr\left(\beta_{1,f}(x^n) \geq 1 - \eta \mid f(x^n) = f(y^n) = 0\right)$$
$$= \Pr\left(\frac{1}{\binom{n}{1}} \sum_{\tilde{y}^n \in \mathcal{S}(x^n, 1) \setminus \{y^n\}} f(\tilde{y}^n) + f(y^n) \geq 1 - \eta \mid f(x^n) = f(y^n) = 0\right)$$
$$= \Pr\left(\frac{1}{n-1} \sum_{\tilde{y}^n \in \mathcal{S}(x^n, 1) \setminus \{y^n\}} f(\tilde{y}^n) \geq \frac{n}{n-1}(1 - \eta)\right)$$
$$= 2^{-(n-1)\left[1 - h\left(\eta + O(n^{-1})\right)\right] + o(n)}$$
$$(23) \qquad = 2^{-n(1 - h(\eta)) + o(n)},$$

where the last transition is since $h(\eta)$ is a smooth function, with bounded derivatives around a neighborhood of any fixed $\eta \in (0, 1)$.

For the second probability on the r.h.s. of (22), if $d_H(x^n, y^n) \geq 3$ then $\mathcal{S}(y^n, 1) \cap \{\{x^n, y^n\} \cup \mathcal{S}(x^n, 1)\} = \phi$ and (20) holds. Next, if $d_H(x^n, y^n) = 1$ then $\mathcal{S}(y^n, 1) \cap$

$\{\{x^n, y^n\} \cup \mathcal{S}(x^n, 1)\} = x^n$. A derivation similar to (23) shows that

$$(24) \qquad \Pr\left(f \in \mathcal{B}_\eta(y^n, 1) \mid f(x^n) = f(y^n) = 0, f \in \mathcal{B}_\eta(x^n, 1)\right) = 2^{-n(1-h(\eta))+o(n)}$$

holds. If $d_H(x^n, y^n) = 2$ then $\mathcal{S}(y^n, 1) \cap \{\{x^n, y^n\} \cup \mathcal{S}(x^n, 1)\}$ contains exactly two points. Again, a derivation similar to (23) (with $n-2$ replacing $n-1$) shows that (24) holds.

The third probability in the r.h.s. of (22) can be trivially upper bounded by 1. Thus,

$$\Pr\left(f \in \mathcal{B}_\eta(x^n) \cap \mathcal{B}_\eta(y^n) \mid f(x^n) = f(y^n) = 0\right) \le 2^{-2n(1-h(\eta))+o(n)}.$$

Evidently, analogous analysis holds for all other three possibilities of the pair $(f(x^n), f(y^n))$,[2] and so

$$(25) \qquad \Pr\left(f \in \mathcal{B}_\eta(x^n) \cap \mathcal{B}_\eta(y^n)\right) \le 2^{-2n(1-h(\eta))+o(n)}.$$

Now, the number of dependent pairs is upper bounded by $2^n \cdot \binom{n}{2\ell}$ since $x^n \sim y^n$ is possible only when $d_H(x^n, y^n) \le 2\ell$. As $\ell = \log n$ was chosen, $\binom{n}{2\ell} \le n^{\log n} = 2^{\log^2 n}$. Then (25) implies that

$$\begin{aligned} \Delta &\le 2^{n+o(n)} \cdot 2^{-2n(1-h(\eta))+o(n)} \\ &= 2^{-n(1-2h(\eta))+o(n)}, \end{aligned}$$

and so $\Delta \to 0$ as $n \to \infty$ exponentially fast, as long as $\eta \in (0, h^{-1}(1/2))$. This concludes the proof. $\qquad \square$

We move on to the proof of Theorem 7.2.

Proof of Theorem 7.2. In this proof we find it more convenient to work with a δ and $\{-1, 1\}$ convention. As the proof of Theorem 7.1, this proof also comprises of a local condition and global analysis. We begin by deriving a necessary condition for a function to be ρ-SP, which is now based only on the value of the function at points of Hamming distance (slightly larger than) $2\eta n$, with $\eta < 1/4$. This condition is tailored to the regime of a fixed δ. We then use a central-limit theorem to show that the probability that this condition is satisfied is close to $1/2$. For global analysis, we consider a subset of the hamming cube of size about $2^{n[1-h(2\eta)]}$ whose minimal Hamming distance is at least ηn. The existence of such a set is guaranteed by the Gilbert-Varshamov bound [Rot06, Th. 4.10]. Since the points in this subset are sufficiently far apart, the event that the local condition holds for one of the points is independent of the corresponding events pertaining to all other points. Thus, the probability of a function to be ρ-SP is not more than about $2^{-2^{n[1-h(2\eta)]}}$.

---

[2]Note that the value of $f(x^n)$ (resp. $f(y^n)$) does not change the asymptotics of the $\Pr(f \in \mathcal{B}_\eta(y^n))$ (resp. $\Pr(f \in \mathcal{B}_\eta(x^n))$).

To prove the required local condition, let $\delta < \eta < 1/4$ be given, and let $\mathcal{D}_\eta(y^n)$ be a punctured Hamming ball of relative radius η around y^n, i.e.,

$$\mathcal{D}_\eta(y^n) := \left\{ z^n \in \{-1,1\}^n : 0 < \frac{1}{n}d_H(z^n, y^n) \leq \eta \right\}.$$

Then, clearly

$$|p(x^n|y^n) \cdot f(x^n)| \leq p(x^n|y^n) = 2^{-n\log\frac{1}{1-\delta}},$$

and by the Chernoff bound (or the method of types [CK11])

$$\left| \sum_{x^n \in \mathcal{D}_\eta^c(y^n)\backslash y^n} p(x^n|y^n) \cdot f(x^n) \right| \leq \sum_{x^n \in \mathcal{D}_\eta^c(y^n)\backslash y^n} p(x^n|y^n)$$

$$\leq \Pr\left(X^n \notin \mathcal{D}_\eta(y^n) \mid Y^n = y^n\right)$$

$$\leq 2^{-nd(\eta||\delta)-\Theta(\log n)}.$$

Focusing on some y^n, we may assume without loss of generality that $f(y^n) = -1$. Then,

$$\mathbb{E}\left(f(X^n) \mid Y^n = y^n\right)$$

$$= \sum_{x^n} p(x^n|y^n) \cdot f(x^n)$$

$$= p(y^n|y^n) \cdot f(y^n) + \sum_{x^n \in \mathcal{D}_\eta(y^n)} p(x^n|y^n) \cdot f(x^n) + \sum_{x^n \in \mathcal{D}_\eta^c(y^n)\backslash y^n} p(x^n|y^n) \cdot f(x^n)$$

$$\geq -2^{-n\log\frac{1}{1-\delta}} + \sum_{x^n \in \mathcal{D}_\eta(y^n)} p(x^n|y^n) \cdot f(x^n) - 2^{-nd(\eta||\delta)-\Theta(\log n)},$$

and thus,

$$\{f \text{ is } \rho\text{-SP at } y^n\} \subseteq$$

$$\left\{ \sum_{x^n \in \mathcal{D}_\eta(y^n)} p(x^n|y^n) \cdot f(x^n) \leq 2^{-n\log\frac{1}{1-\delta}} + 2^{-nd(\eta||\delta)-\Theta(\log n)} \right\} := \mathcal{A}_\eta(y^n).$$

We next evaluate the probability that the necessary condition is satisfied when f is chosen uniformly at random over all Boolean functions on $\{-1,1\}^n$. Specifically, we use the Berry-Esseen central-limit theorem [Fel71, Chapter XVI.5, Theorem 2] to bound $\Pr(\mathcal{A}_\eta(y^n))$. To that end, we note that

$$\mathbb{E}\left(\sum_{x^n \in \mathcal{D}_\eta(y^n)} p(x^n|y^n) \cdot f(x^n) \right) = 0,$$

and that by the method of types [CK11]

$$\mathbb{E}\left(\sum_{x^n \in \mathcal{D}_\eta(y^n)} p(x^n|y^n) \cdot f(x^n)\right)^2 = \sum_{x^n \in \mathcal{D}_\eta(y^n)} p^2(x^n|y^n)$$

$$= \sum_{\ell=1}^{\lfloor \eta n \rfloor} \sum_{x^n : d_H(x^n,y^n)=\ell} 2^{-2n[h(\ell/n)+d(\ell/n||\delta)]}$$

(26)
$$= 2^{-2n\cdot\min_{0\leq\zeta\leq\eta}[h(\zeta)+2d(\zeta||\delta)]-\Theta(\log n)}$$

$$= 2^{-n\cdot\log\frac{1}{\delta^2+(1-\delta)^2}-\Theta(\log n)}$$

where $\zeta := d/n$, and the minimum in (26) is attained for $\zeta = \frac{\delta^2}{\delta^2+(1-\delta)^2}$ (which satisfies $\zeta \leq \delta < \eta$). Similarly, we note that

$$\gamma_n := \sum_{x^n \in \mathcal{D}_\eta(y^n)} \mathbb{E}\left|p(x^n|y^n) \cdot f(x^n)\right|^3$$

$$= 2^{-n\cdot\min_{0\leq\zeta\leq\eta}[2h(\zeta)+3d(\zeta||\delta)]-\Theta(\log n)},$$

where clearly γ_n decreases exponentially for any $\delta \in (0, 1/2)$. Consequently, the Berry-Esseen central-limit theorem implies that there exists a universal constant c such that

(27)
$$\Pr\left(\mathcal{A}_\eta(y^n)\right) \leq 1 - Q\left(\frac{2^{-n\log\frac{1}{1-\delta}} + 2^{-nd(\eta||\delta)-\Theta(\log n)}}{2^{-n\cdot\frac{1}{2}\log\frac{1}{\delta^2+(1-\delta)^2}-\Theta(\log n)}}\right) + c\gamma_n$$

(28)
$$\leq \frac{1}{2} + o(1).$$

where in (27) $Q(\cdot)$ is the tail distribution function of the standard normal distribution, and (28) is satisfied whenever $\eta > \eta_\delta$. This completes the analysis of the local necessary condition.

We next move on to global analysis. By the Gilbert-Varshamov bound [Rot06, Th. 4.10], there exists a set (also known as an *error-correcting code*) $\mathcal{C}_n \subset \{-1, 1\}^n$ such that

$$|\mathcal{C}_n| \geq 2^{n[1-h(2\eta)]-o(n)}$$

and $\mathcal{D}_\eta(x^n) \cap \mathcal{D}_\eta(y^n) = \phi$ for all $x^n, y^n \in \mathcal{C}_n$. Consequently,

$$\Pr\left(f \text{ is } \rho\text{-SP}\right) \leq \Pr\left(\bigcap_{y^n \in \mathcal{C}_n} f \text{ is } \rho\text{-SP at } y^n\right)$$

$$\leq \Pr\left(\bigcap_{y^n \in \mathcal{C}_n} \{f \in \mathcal{A}_\eta^c(y^n)\}\right)$$

$$= \prod_{y^n \in \mathcal{C}_n} \Pr\left(f \in \mathcal{A}_\eta(y^n)\right)$$

$$\leq \left(\frac{1}{2} + o(1)\right)^{|\mathcal{C}_n|}.$$

The proof is completed since $|\mathcal{C}_n|$ increases exponentially for $\eta < 1/4$. $\qquad\square$

Remark 7.3. It is evident that the proof of Theorem 7.2 also holds for any sequence of $\{\delta_n\}$ such that $\delta_n = \omega(\frac{1}{\sqrt{n}})$ and $\overline{\delta} := \limsup_{n\to\infty} \delta_n < \delta_{\max}$. Indeed, (28) holds in this case too, as long as we choose η to be $\eta_{\overline{\delta}}$.

8. OPEN PROBLEMS

We have introduced the notion of self-predictability for Boolean functions. There are many interesting questions left open; below is far from an exhaustive list.

We know that the characters, Majority and a few other LTFs (found numerically) are USP, and we can create many other USP functions from them. However, we still lack a clear understanding of what makes a function USP.

Problem 8.1. Characterize the family of USP functions. Specifically, how many USP functions are there?

More specifically, we ask:

Problem 8.2. Is there a finite set of USP functions and a finite set of SP-preserving operations that span all USP functions?

Adding symmetry to the mix, we conjecture the following.

Conjecture 8.3. *The only symmetric USP functions are Majority and the largest character. In particular, Majority is the only monotone and symmetric USP function.*

We have seen that LCSP functions are WST, but not vice versa.

Problem 8.4. Find a simple condition guaranteeing that a WST function is LCSP (resp. USP).

We say that a function f is *monotonically SP* if there exists ρ_0 such that f is ρ-SP for $\rho > \rho_0$ and not ρ-SP for $\rho < \rho_0$. We have seen that there exist (balanced) functions that are not monotonically SP.

Problem 8.5. Characterize the family of monotonically SP functions.

We have bounded the ratio between the strongest and second strongest coefficient of an LCSP LTF. This is quite weak: Let $r_n(\mathcal{F})$ be the minimum number such that any LTF in the family $\mathcal{F}$ admits a representation in which the ratio between the maximal coefficient and minimal coefficient (in absolute values) is at most $r_n(\mathcal{F})$. It is known

in general, (see [BHPS10, Theorem 2] and references therein) that $2^{-n(2-o(1))} \cdot n^{n/2} \leq r_n(\mathcal{F}) \leq 2^{n-1} \cdot (n+1)^{(n+1)/2}$ if $\mathcal{F}$ is the family of all LTFs. It is interesting to ask whether $r_n(\mathcal{F})$ becomes much smaller under self-predictability.

Problem 8.6. Characterize $r_n(\mathcal{F})$ when $\mathcal{F}$ is the family of LCSP LTFs (resp. USP LTFs).

Let $\mathcal{G}_{\rho,n}$ be a directed graph over the set of all Boolean functions with n variables, where we draw a directed edge from every function f to its optimal predictor $\operatorname{sgn} T_\rho f$ (unless they coincide). To avoid ambiguities, we can set $\operatorname{sgn} T_\rho f$ equal to f whenever $T_\rho f$ is exactly zero. It is easy to see that the number of ρ-SP functions is upper bounded by the number of weakly connected components of $\mathcal{G}_{\rho,n}$, namely the connected components of the associated undirected graph obtained by removing the direction of the edges. In fact, we conjecture that these quantities are exactly equal, or equivalently:

Conjecture 8.7. $\mathcal{G}_{\rho,n}$ *contains no cycles.*

Note that if the above conjecture holds, then a simple way to arrive at a ρ-SP function is to start with some function f and repeatedly apply the $\operatorname{sgn} T_\rho$ operator; this procedure will terminate at a ρ-SP function in finite time. In fact, simulations indicate that this convergence happens very quickly, which may hint that the weakly connected components of $\mathcal{G}_{\rho,n}$ have small depth.

ACKNOWLEDGMENTS

REFERENCES

[AS04] N. Alon and J. H. Spencer, *The probabilistic method*, John Wiley & Sons, 2004.

[BHPS10] L. Babai, K. A. Hansen, V. V. Podolskii, and X. Sun, *Weights of exact threshold functions*, MFCS, Springer, 2010, pp. 66–77.

[Bon70] A. Bonami, *Étude des coefficients de Fourier des fonctions de lp(g)*, Ann. Inst. Fourier (Grenoble) **20** (1970), no. 2, 335–402.

[Bru90] J. Bruck, *Harmonic analysis of polynomial threshold functions*, SIAM Journal on Discrete Mathematics **3** (1990), no. 2, 168–177.

[Cho61] C.-K. Chow, *On the characterization of threshold functions*, Switching Circuit Theory and Logical Design, 1961. SWCT 1961. Proceedings of the Second Annual Symposium on, IEEE, 1961, pp. 34–38.

[CK11] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless systems*, Cambridge University Press, 2011.

[Fel71]   W. Feller, *An introduction to probability theory and its applications*, vol. 2, John Wiley & Sons, New York, 1971.

[GM99]   N. K. Govil and R. N. Mohapatra, *Markov and Bernstein type inequalities for polynomials*, Journal of Inequalities and Applications **3** (1999), no. 4, 349–387.

[Gro75]   L. Gross, *Logarithmic Sobolev inequalities*, American Journal of Mathematics **97** (1975), no. 4, 1061–1083.

[Haa81]   U. Haagerup, *The best constants in the Khintchine inequality*, Studia Mathematica **70** (1981), no. 3, 231–283.

[Kel10]   N. Keller, *On the probability of a rational outcome for generalized social welfare functions on three alternatives*, Journal of Combinatorial Theory Series A **117** (2010), no. 4, 389–410.

[KKM16]   G. Kalai, N. Keller, and E. Mossel, *On the correlation of increasing families*, Journal of Combinatorial Theory, Series A **144** (2016), 250–276.

[O'D14]   R. O'Donnell, *Analysis of Boolean functions*, Cambridge University Press, 2014.

[OS11]   R. O'Donnell and R. A. Servedio, *The Chow parameters problem*, SIAM Journal on Computing **40** (2011), no. 1, 165–199.

[Rot06]   R. Roth, *Introduction to coding theory*, Cambridge University Press, 2006.

[STlV17]   A. Shpilka, A. Tal, and B. lee Volk, *On the structure of Boolean functions with small spectral norm*, computational complexity **26** (2017), no. 1, 229–273.

[Tan61]   M. Tannenbaum, *The establishment of a unique representation for a linearly separable function*, Lockheed Missilesand Space Co., Sunnyvale, Calif., Threshold Switching Techniques Note **20** (1961), 1–5.