A Bivariate Invariance Principle

Alexander Mariona*, Homa Esfahanizadeh*, Rafael G. L. D’Oliveira†, and Muriel Médard*

* Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139
† School of Mathematical and Statistical Sciences, Clemson University, Clemson, SC 29634

Email: amariona@mit.edu, homaesf@mit.edu, rdolive@clemson.edu, medard@mit.edu

Abstract—A notable result from analysis of Boolean functions is the Basic Invariance Principle (BIP), a quantitative nonlinear generalization of the Central Limit Theorem for multilinear polynomials. We present a generalization of the BIP for bivariate multilinear polynomials, i.e., polynomials over two $n$-length sequences of random variables. This bivariate invariance principle arises from an iterative application of the BIP to bound the error in replacing each of the two input sequences. In order to prove this invariance principle, we first derive a version of the BIP for random multilinear polynomials, i.e., polynomials whose coefficients are random variables. As a benchmark, we also state a naive bivariate invariance principle which treats the two input sequences as one and directly applies the BIP. Neither principle is universally stronger than the other, but we do show that for a notable class of bivariate functions, which we term separable functions, our subtler principle is exponentially tighter than the naive benchmark.

Index Terms—Basic Invariance Principle, Boolean functions, functional approximation

An extended version of this paper including all proofs can be read at https://arxiv.org/abs/2208.04977.

I. INTRODUCTION

Boolean functions are ubiquitous in the fields of complexity theory [1], [2], cryptography [3], [4], social choice theory [5], [6], and digital electronics [7], [8]. One particularly significant result from the field of analysis of Boolean functions is the Basic Invariance Principle (BIP) [9]. The BIP is a nonlinear generalization of the Berry-Esseen Theorem [10], [11], which is in turn a quantitative version of the Central Limit Theorem. The Berry-Esseen Theorem provides an explicit bound on the difference between the distribution of a finite sum of independent random variables and a standard Gaussian distribution. The BIP, given a multilinear polynomial and two differently distributed sequences of independent random variables $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$, bounds the expected difference between $f(X)$ and $f(Y)$. This difference can be interpreted as the expected error incurred by approximating $f(X)$ as $f(Y)$.

In order for the bound given by the BIP to be close to 0, the function under consideration must have relatively low influences. The influence of a coordinate on a Boolean function quantifies how sensitive the output is to a change in that particular input coordinate, and the same concept can be generalized to multilinear polynomials. The notion of influence originated in social choice and voting theory [12].

Qualitatively, the BIP states that low-influence functions are invariant to the distribution of the input sequence. One notable application of the BIP is to “replace bits by Gaussians:” whether the input is a sequence of uniform random bits or a sequence of standard Gaussians, the expected output of a low-influence function does not change too much.

One natural generalization of the BIP would be an invariance principle which treats functions of two sequences of random variables. Such bivariate functions open up new options and ideas in applications involving two distinct data sources, such as in multi-party communication networks, e.g., [13], [14]. In the context of these models, some functions are inherently bivariate, even if they could be equivalently written as univariate functions. In those cases, a bivariate generalization of the BIP may achieve a tighter bound by exploiting the bivariate structure of the function.

We present one such invariance principle which follows from iteratively applying the BIP to bound the error in replacing the first input sequence and then again to bound the error in replacing the second. In order to do so, we treat the bivariate function as a univariate function with random coefficients which are determined by the input sequence that is not being replaced at a given step. To this end, we propose a variation of the BIP which can be applied to such random functions. For the sake of comparison, we also derive a naive bivariate invariance principle directly by treating the two input sequences as a single sequence, effectively viewing the bivariate function as univariate. We refer to our subtler invariance principle as BVIP-1 and to this naive benchmark as BVIP-2. Neither principle is universally stronger than the other, but we do offer one notable example of a family of functions for which BVIP-1 is exponentially tighter: functions of the form $F(x,y) = f(x) + g(y) + h(xy)$, which we term separable functions. These functions are particularly interesting because they generalize many different notions of noise that arise in communication channels, including simple models like the binary symmetric channel [14], [15].

The remainder of this paper is organized as follows. In Section II, we summarize concepts and review key results from analysis of Boolean functions. In Section III, we consider multilinear random polynomials and propose a version of the BIP for those random functions in anticipation of Section IV, in which we prove and prove BVIP-1. In Section V, we compare BVIP-1 to the naive benchmark of BVIP-2, present corollary invariance principles for the special case of separable functions, and offer concluding thoughts.

This work was supported by MIT Welcome Trust Fellowship 2389724, Lincoln Lab Grant 6944494, MIT Portugal Grant 6942770, and NFS-CNS Grant 6944400.
II. PRELIMINARIES

We denote random variables with uppercase letters, e.g., $X$. We denote vectors (often referred to as sequences in our context) with bold-faced letters, e.g., $\mathbf{x}$. Accordingly, vectors of random variables are denoted with uppercase bold-faced letters, e.g., $\mathbf{X}$. We denote the coordinates (or elements) of vectors with indexed letters, e.g., $x_i$. We sometimes specify the coordinates of a vector like $x = (x_1, x_2, \ldots, x_n)$. Multiplication of two vectors is performed elementwise and results in a new vector, i.e., $\mathbf{xy} = (x_1y_1, x_2y_2, \ldots, x_ny_n)$. We denote the set containing the element $i$ with $S \ni i$. We denote the set $\{1, 2, \ldots, n\}$ with $[n]$ and its power set with $2^n$.

A. Results from Analysis of Boolean Functions

We begin with Boolean functions $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ and real-valued Boolean functions $f : \{-1, 1\}^n \rightarrow \mathbb{R}$. All results in the following sections in fact hold for general multilinear polynomials $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (the domain is $\mathbb{R}^n \times \mathbb{R}^n = \mathbb{R}^{2n}$ in the bivariate case), but because many of the key tools are defined in the context of Boolean functions, we briefly discuss those functions here before generalizing. All definitions and results in this section other than Definition 3 are from [16].

Theorem 1. Every Boolean function $f : \{-1, 1\}^n \rightarrow \mathbb{R}$ can be uniquely expressed as an $n$-variate multilinear polynomial,

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

This expression is called the Fourier expansion of $f$ and is determined by the Fourier coefficients of $f$ on $S$ which are given by the function $\hat{f} : 2^n \rightarrow \mathbb{R}$. Collectively, the coefficients of $f$ are referred to as the Fourier spectrum of $f$. When we refer to the degree of a Boolean function, we are referring to the degree of its Fourier expansion. Since every such expansion is multilinear, the degree $k$ of a Boolean function $f$ (or of any multilinear polynomial $f$) is

$$ k = \max_{f(S) \neq 0} |S|. $$

An important property of a Boolean function is the influence of each coordinate of the input on the output of the function. The influence of a coordinate quantifies how likely a particular coordinate is to be pivotal. A coordinate $i$ is pivotal for a particular input $x$ if negating $x_i$ negates the output $f(x)$.

Definition 1. The influence of a coordinate $i$ on a function $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ is defined to be the probability that $i$ is pivotal for a random input drawn uniformly:

$$ \text{Inf}_i[f] = \mathbb{E}_{x \sim \{-1, 1\}^n} [f(x) \neq f(x^{\oplus i})], $$

where $x^{\oplus i} = (x_1, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_n)$.

Informally, if we consider $f$ to be a voting rule in a two-party election, the influence of the $i$th coordinate can be thought of as the “influence” or “power” of the $i$th voter. The influences of a real-valued Boolean function can be defined in a more analytical fashion, but with a very similar meaning. Such an approach leads to a relation between the influences and the Fourier spectrum which we treat as a definition.

Definition 2. For $f : \{-1, 1\}^n \rightarrow \mathbb{R}$ and $i \in [n]$, the influence of coordinate $i$ on $f$ is

$$ \text{Inf}_i[f] = \sum_{S \ni i} \hat{f}(S)^2. $$

We also use Definition 2 for general multilinear polynomials, a choice which is justified by Lemma 2 below.¹

We now present a few statements in anticipation of the BIP. First, the BIP only holds for sequences of random variables with well-behaved distributions. In particular, we make the following assumption on each random variable in the two sequences with which we are concerned.²

Assumption 1. The random variable $X_i$ satisfies $\mathbb{E}[X_i] = 0$, $\mathbb{E}[X_i^2] = 1$, $\mathbb{E}[X_i^3] = 0$, and $\mathbb{E}[X_i^4] \leq 9 \mathbb{E}[X_i^2]^2$. Let $f$ be a multilinear polynomial of degree at most $k$. Then

$$ \mathbb{E}[f(X_i^4)] \leq 9^k \cdot \mathbb{E}[f(X_i^2)^2]. $$

Lemma 1 (Bonami Lemma). Let $X = (X_1, \ldots, X_n)$ be a sequence of independent but not necessarily identically distributed random variables satisfying $\mathbb{E}[X_i] = 0$, $\mathbb{E}[X_i^2] = 1$, $\mathbb{E}[X_i^3] = 0$, and $\mathbb{E}[X_i^4] \leq 9 \mathbb{E}[X_i^2]^2$. Let $f$ be a multilinear polynomial of degree at most $k$. Then

$$ \mathbb{E}[f(X_i^4)] \leq 9^k \cdot \mathbb{E}[f(X_i^2)^2]. $$

Lemma 2. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be an $n$-variate multilinear polynomial over the sequence of indeterminates $x = (x_1, \ldots, x_n)$,

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

When considering a sequence of independent random variables $X = (X_1, \ldots, X_n)$ with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i^2] = 1$, the parity functions $\chi_S = \prod_{i \in S} X_i$ are orthonormal, and hence

$$ \mathbb{E}[f(X^2)] = \sum_{S \subseteq [n]} \hat{f}(S)^2. $$

This leads us to the formal statement of the BIP.

Theorem 2 (BIP). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be an $n$-variate multilinear polynomial of degree at most $k$.

$$ \mathbb{E}[f(X_i^4)] \leq 9^k \cdot \mathbb{E}[f(X_i^2)^2]. $$

¹Lemma 2 states that Parseval’s theorem holds for multilinear polynomials which are applied to sequences satisfying Assumption 1. It is because of this that we are justified in using Definition 2. See [16, ch. 8.2] for more detail.

²A slightly different form of the BIP holds for a looser set of assumptions. Assumption 1 is a simpler hypothesis which keeps the bounds tidy and will suffice for our purposes. See [9, sec. 3.3] for more detail.
The function $\psi$ used in the BIP is called a test function or a distinguisher, and is used to specify a particular notion of “closeness” between two random variables. A natural measure is cdf-closeness, which is used in the Berry-Esseen Theorem. Two random variables $X$ and $Y$ are cdf-close if $\Pr\{X \leq u\} \approx \Pr\{Y \leq u\}$ for all $u \in \mathbb{R}$. Equivalently, two random variables are cdf-close if $E[\psi(X)] \approx E[\psi(Y)]$ with $\psi(s) = 1_{s \leq u}$ for all $u \in \mathbb{R}$. The BIP is powerful enough to give bounds on cdf-closeness and many other notions of closeness. \(^3\)

### B. Bivariate Functions

Finally, we specify the class of functions which we will refer to throughout this paper simply as bivariate functions.

**Definition 3.** An $n$-bivariate multilinear polynomial function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ over the sequences of indeterminates $x_1 = (x_{11}, \ldots, x_{1n})$ and $x_2 = (x_{21}, \ldots, x_{2n})$ is a function of the form

$$f(x_1, x_2) = \sum_{S_1, S_2 \subseteq [n]} \hat{f}(S_1, S_2) \prod_{i \in S_1} x_{1i} \prod_{j \in S_2} x_{2j}.$$  

The form given in Definition 3 suggests that $\hat{f}(S_1, S_2)$ is Fourier coefficient. Indeed, if we consider $f$ to instead be a function of the concatenated sequence $x = (x_1||x_2)$, then $\hat{f}(S_1, S_2)$ is the Fourier coefficient on the set $S_1 \cup S_2^+$, where $S_2^+ = \{i + n : i \in S_2\}$. Nonetheless, we will not consider any subtleties of Fourier theory for bivariate functions and we do not make any claims about any of the classic Fourier identities in this bivariate basis.

### III. RANDOM FUNCTIONS

In anticipation of BVIP-1, we introduce in this section the concept of random multilinear polynomials and prove a version of the BIP for these functions.

**Definition 4.** A random $n$-variate multilinear polynomial $f_Z : \mathbb{R}^n \to \mathbb{R}$ is a polynomial multilinear over the sequence of indeterminates $x = (x_1, \ldots, x_n)$ whose coefficients $f_Z(S) \in \mathbb{R}$ are random variables which are wholly determined by the random variable $Z$:

$$f_Z(x) = \sum_{S \subseteq [n]} \hat{f}_Z(S) \prod_{i \in S} x_i.$$  

The influence of coordinate $i$ on $f_Z$ is a random variable which is defined to be

$$\text{Inf}_i[f_Z] = \sum_{S \ni i} \hat{f}_Z(S)^2.$$  

We think of $Z$ as the random variable which controls $f_Z$ or, alternatively, which describes the randomness of $f_Z$. We will sometimes refer to random multilinear polynomials simply as random functions. Such functions will always be univariate.

In pursuit of a version of the BIP for random functions, we start with appropriate corollaries of Lemma 1 and Lemma 2.

**Corollary 1.** Let $X = (X_1, \ldots, X_n)$ be a sequence of independent but not necessarily identically distributed random variables satisfying the requirement that $E[X_1] = E[X_2] = 0$ and $E[X_i^4] \leq 9 E[X_i^2]^2$. Let $f_Z$ be a random multilinear polynomial of degree at most $k$. Then

$$E_{X,Z}[f_Z(X)^4] \leq 9^k \cdot E_{X,Z}[f_Z(X)^2]^2.$$  

**Corollary 2.** Let $f_Z$ be a random $n$-variate multilinear polynomial. When considering a sequence of independent random variables $X = (X_1, \ldots, X_n)$ satisfying $E[X_i] = 0$ and $E[X_i^2] = 1$, the parity functions $\chi_S = \prod_{i \in S} X_i$ are orthonormal, and hence

$$E_{X,Z}[f_Z(X)^2] = E_{Z}[\sum_{S \subseteq [n]} \hat{f}_Z(S)^2].$$  

Our BIP for random functions is identical in spirit to Theorem 2, but the resulting upper bound is in terms of the expected influences of the given function.

**Theorem 3 (BIP for Random Functions).** Let $f_Z$ be a random $n$-variate multilinear polynomial of degree at most $k$. Let $X$ and $Y$ be $n$-length sequences of independent random variables satisfying Assumption 1. Let $\psi : \mathbb{R} \to \mathbb{R}$ be $C^4$. Then

$$|E_Z[\psi(f_Z(X))] - E_Z[\psi(f_Z(Y))]| \leq \frac{C}{12} 9^k \sum_{i=1}^n E[Z^2]\text{Inf}_i[f_Z]^2.$$  

For the sake of brevity and due to its similarity to the proof of the BIP given in [16, ch. 11.6], we omit the full proof of Theorem 3. The only differences are that all expectations are taken over $Z$ and that Corollary 1 and Corollary 2 are used in place of Lemma 1 and Lemma 2 respectively.

### IV. A BIVARIATE INVARIANCE PRINCIPLE

We now present BVIP-1. To derive it, we iteratively apply the BIP to replace each input sequence in turn. The first step in this process is to treat the input sequence which is not currently being replaced as a random parameter of the function, allowing us to view the bivariate function as a random univariate function. We can then use the BIP for random functions to bound the error incurred by this replacement.

**Theorem 4 (BVIP-1).** Let $f$ be an $n$-bivariate multilinear polynomial in which each term includes at most $k$ elements from each input sequence:

$$f(x_1, x_2) = \sum_{S_1, S_2 \subseteq [n]} \hat{f}(S_1, S_2) \prod_{i \in S_1} x_{1i} \prod_{j \in S_2} x_{2j},$$  

where $\hat{f}(S_1, S_2) = 0$ if $|S_1| > k$ or $|S_2| > k$. Let $X_1, X_2, Y_1,$ and $Y_2$ be $n$-length sequences of independent random variables satisfying Assumption 1. Assume $\psi : \mathbb{R} \to \mathbb{R}$ is $C^4$ with $\|\psi^{(4)}\|_\infty \leq C$. Then

$$|E_X - E_Y| \leq \frac{C}{12} 9^k \cdot \sum_{t=1}^n (\hat{S}_{1,t}^2 + \hat{S}_{2,t}^2),$$  

(1)
where

\[ E_X = \mathbb{E}_{X_1, X_2}[\psi(f(X_1, X_2))] \]

\[ E_Y = \mathbb{E}_{Y_1, Y_2}[\psi(f(Y_1, Y_2))] \]

\[ \tilde{\Sigma}_{1,t} = \sum_{S_1 \not\subseteq \Theta} [T_2(S_1)] \sum_{S_2 \subseteq T_2(S_1)} \hat{f}(S_1, S_2)^2 \]

\[ \tilde{\Sigma}_{2,t} = \sum_{S_2 \not\subseteq \Theta} [T_1(S_2)] \sum_{S_1 \subseteq T_1(S_2)} \hat{f}(S_1, S_2)^2, \]

and \( T_2(S_1) \) and \( T_1(S_2) \) are the sets

\[ T_2(S_1) = \{ S_2 \subseteq [n] : |S_2| \leq k, \hat{f}(S_1, S_2) \neq 0 \} \]

\[ T_1(S_2) = \{ S_1 \subseteq [n] : |S_1| \leq k, \hat{f}(S_1, S_2) \neq 0 \}. \]

\textbf{Proof.} As described above, our strategy is to define random functions \( f_{X_2} \) and \( f_{Y_1} \) such that \( f_{X_2}(X_1) = f(X_1, X_2) \) and \( f_{Y_1}(X_2) = f(Y_1, X_2) \). Applying Theorem 3 to \( f_{X_2} \) bounds the error incurred by replacing \( X_1 \) with \( Y_1 \). An application of the same theorem to \( f_{Y_1} \) bounds the error in replacing \( X_2 \) and \( Y_2 \). Computing the expected influences of the random functions and using the triangle inequality will complete the proof.

We begin by constructing the desired random functions. Let \( f_{X_2} : \mathbb{R}^n \to \mathbb{R} \) be defined as

\[ f_{X_2}(t) = f(t, X_2) = \sum_{S_1 \subseteq [n]} \left( \sum_{S_2 \subseteq \Theta} \hat{f}(S_1, S_2) \prod_{j \in S_2} X_{2,j} \right) \prod_{i \in S_1} t_i \]

Similarly, let \( f_{Y_1} : \mathbb{R}^n \to \mathbb{R} \) be defined as

\[ f_{Y_1}(t) = f(t, Y_1) = \sum_{S_1 \subseteq [n]} \left( \sum_{S_2 \subseteq \Theta} \hat{f}(S_1, S_2) \prod_{j \in S_2} Y_{1,j} \right) \prod_{i \in S_2} t_i \]

Note that both \( f_{X_2} \) and \( f_{Y_1} \) are of degree at most \( k \) and that \( f_{X_2}(Y_1) = f_{Y_1}(X_2) \). From the definitions of \( E_X \) and \( E_Y \),

\[ E_X = \mathbb{E}_{X_1, X_2}[\psi(f_{X_2}(X_1))] \]

\[ E_Y = \mathbb{E}_{Y_1, Y_2}[\psi(f_{Y_1}(Y_2))] \]

By analogy, let

\[ E_{X_Y} = \mathbb{E}_{Y_1, X_2}[\psi(f(Y_1, X_2))] \]

\[ = \mathbb{E}_{Y_1, X_2}[\psi(f_{X_2}(Y_1))] \]

\[ = \mathbb{E}_{Y_1, X_2}[\psi(f_{Y_1}(X_2))]. \]

We upper bound the quantity of interest as

\[ |E_X - E_Y| \leq |E_X - E_{X_Y}| + |E_{X_Y} - E_Y|. \]

Applying Theorem 3 to each term on the right-hand side of (2) yields

\[ |E_X - E_{X_Y}| \leq \frac{C}{12} \cdot 9^k \cdot \sum_{t=1}^{n} \mathbb{E}_{X_2}[\text{Inf}_t[f_{X_2}]]^2 \]  

(3)

\[ |E_{X_Y} - E_Y| \leq \frac{C}{12} \cdot 9^k \cdot \sum_{t=1}^{n} \mathbb{E}_{Y_1}[\text{Inf}_t[f_{Y_1}]]^2. \]  

(4)

All that remains is to bound the expected influences of \( f_{X_2} \) and \( f_{Y_1} \). We handle the case of \( f_{X_2} \) explicitly, with the argument for \( f_{Y_1} \) being identical. For convenience, define

\[ \sigma_2(S_1) = \sum_{S_2 \subseteq T_2(S_1)} \hat{f}(S_1, S_2)^2. \]

We have

\[ \mathbb{E}_{X_2}[\text{Inf}_t[f_{X_2}]] = \mathbb{E}_{X_2}\left[ \sum_{S_1 \not\subseteq \Theta} \hat{f}(S_1, S_2)^2 \right] \]

\[ = \mathbb{E}_{X_2}\left[ \sum_{S_1 \not\subseteq \Theta} \left( \sum_{S_2 \subseteq \Theta} \hat{f}(S_1, S_2) \prod_{j \in S_2} X_{2,j} \right)^2 \right] \]

\[ = \sum_{S_1 \not\subseteq \Theta} \left( \sum_{S_2 \subseteq T_2(S_1)} \hat{f}(S_1, S_2) \prod_{j \in S_2} X_{2,j} \right)^2 \]

(5)

\[ \leq \sum_{S_1 \not\subseteq \Theta} \sum_{S_2 \subseteq T_2(S_1)} \hat{f}(S_1, S_2)^2 \prod_{j \in S_2} X_{2,j}^2 \]

\[ \leq \sum_{S_1 \not\subseteq \Theta} \sigma_2(S_1) \sum_{S_2 \subseteq T_2(S_1)} \prod_{j \in S_2} X_{2,j}^2 \]

(6)

\[ \leq \sum_{S_1 \not\subseteq \Theta} \sigma_2(S_1) \sum_{S_2 \subseteq T_2(S_1)} \prod_{j \in S_2} X_{2,j}^2 \]

(7)

\[ = \sum_{S_1 \not\subseteq \Theta} \left( \sum_{S_2 \subseteq T_2(S_1)} \prod_{j \in S_2} X_{2,j}^2 \right) \]

(8)

\[ = \sum_{S_1 \not\subseteq \Theta} \left( \sum_{S_2 \subseteq T_2(S_1)} \prod_{j \in S_2} X_{2,j}^2 \right) \]

(9)

where (5) follows from linearity of expectation and the fact that for a given \( S_1 \), we have \( \hat{f}(S_1, S_2) \neq 0 \) only if \( S_2 \subseteq T_2(S_2) \); (6) follows from the Cauchy-Schwarz inequality; (7) again follows from linearity of expectation; (8) follows from the assumption that the elements of \( X_2 \) are independent; and (9) follows from the assumption that \( \mathbb{E}[X_{2,j}^2] = 1 \) for all \( j \in [n] \). The same argument applied to \( f_{Y_1} \) gives

\[ \mathbb{E}_{Y_1}[\text{Inf}_t[f_{Y_1}]] \leq \tilde{\Sigma}_{2,t}. \]  

(11)

Substituting (10) and (11) into (3) and (4) respectively yields

\[ |E_X - E_{X_Y}| \leq \frac{C}{12} \cdot 9^k \cdot \sum_{t=1}^{n} \tilde{\Sigma}_{2,t} \]

\[ |E_{X_Y} - E_Y| \leq \frac{C}{12} \cdot 9^k \cdot \sum_{t=1}^{n} \tilde{\Sigma}_{2,t}, \]

which, combined with (2), completes the proof. \( \square \)
V. DISCUSSION AND CONCLUSION

As a baseline against which we can compare the bound of Theorem 4, we state as a corollary the bound which the BIP yields when we treat a bivariate function as a univariate function. Such a univariate interpretation takes as input the concatenation of the two sequences which are the inputs to the original bivariate function. We refer to this naive bivariate invariance principle as BVIP-2.

Corollary 3 (BVIP-2). Let $f$ be an $n$-bivariate multilinear polynomial in which each term includes at most $k$ elements from each input sequence:

$$f(x_1, x_2) = \sum_{S_1, S_2 \subseteq [n]} \hat{f}(S_1, S_2) \prod_{i \in S_1} x_{1,i} \prod_{j \in S_2} x_{2,j},$$

where $\hat{f}(S_1, S_2) = 0$ if $|S_1| > k$ or $|S_2| > k$. Let $X_1$, $X_2$, $Y_1$, and $Y_2$ be $n$-length sequences of independent random variables satisfying Assumption 1. Assume $\psi : \mathbb{R} \to \mathbb{R}$ is $C^4$ with $\|\psi^{(m)}\|_{\infty} \leq C$. Then

$$|E_X - E_Y| \leq \frac{C}{12} \cdot 9^{2k} \cdot \sum_{t=1}^{n} (\Sigma_{1,t}^2 + \Sigma_{2,t}^2),$$

where $E_X$, $E_Y$, $T_2(S_1)$ and $T_1(S_2)$ are as defined in Theorem 4, and

$$\Sigma_{1,t} = \sum_{S_1 \neq S_2 \in T_2(S_1)} \sum_{S_1 \neq S_2 \in T_2(S_1)} \hat{f}(S_1, S_2)^2$$

$$\Sigma_{2,t} = \sum_{S_2 \neq S_1 \in T_2(S_2)} \sum_{S_2 \neq S_1 \in T_2(S_2)} \hat{f}(S_1, S_2)^2.$$

Note that in the context of BVIP-2, $k$ is not strictly speaking the degree of $f$, as it is in the BIP. Indeed, the degree of $f$ can be as large as $2k$ here, and as such, the bound incurs a factor of $9^{2k}$ directly from the BIP.

Comparing the bounds of BVIP-1 and BVIP-2 given in (1) and (12) respectively, the main differences are a factor of $9^{k}$ versus $9^{2k}$ and the quantity $\Sigma_{1,t}$ versus $\Sigma_{1,t}$ (for $i \in \{1, 2\}$), which we recall are defined (for $i = 1$) as

$$\Sigma_{1,t} = \sum_{S_1 \neq S_2 \in T_2(S_1)} \sum_{S_2 \in T_2(S_1)} \hat{f}(S_1, S_2)^2$$

$$\Sigma_{1,t} = \sum_{S_1 \neq S_2 \in T_2(S_1)} \sum_{S_2 \in T_2(S_1)} \hat{f}(S_1, S_2)^2.$$

Thus, BVIP-1 trades a factor of $9^{k}$ compared to BVIP-2 in exchange for counting $|T_2(S_1)|$ for each $S_1$ (and likewise $|T_2(S_2)|$ for each $S_2$). We can conceptualize $|T_2(S_1)|$ and $|T_1(S_2)|$ as measuring the “strength” of the interaction in $f$ between the inputs $X_1$ and $X_2$. If those cardinalities are large, then there are many terms of $f$ in which some coordinates of $X_2$ are multiplied with the coordinates of $X_1$. Note that $|T_2(S_1)|$ and $|T_2(S_1)|$ both arise from applying the Cauchy-Schwarz inequality, as in (6), and are hence upper bounds on this interaction strength.

The question of whether BVIP-1 outperforms BVIP-2 for a particular $f$ is thus a question of whether the interaction between $X_1$ and $X_2$ is “small enough” to beat the extra factor of $9^k$ incurred by BVIP-2. As a concrete example of a family of functions for which BVIP-1 is always tighter than BVIP-2, consider separable bivariate functions.

Definition 5. An $n$-bivariate multilinear polynomial function $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is separable into $f$, $g$, and $h$ if it can be written in terms of $n$-variate multilinear polynomials $f, g, h : \mathbb{R}^n \to \mathbb{R}$ like

$$F(x_1, x_2) = f(x_1) + g(x_2) + h(x_1, x_2).$$

For separable functions, the bounds of both bivariate invariance principles can be cleanly expressed in terms of the influences of $f$, $g$, and $h$, resulting in a form which is very close to that of the BIP.

Corollary 4 (Separable BVIP-1). Let $F$ be an $n$-bivariate multilinear polynomial which is separable into $f$, $g$, and $h$, each of which is of degree at most $k$. Let $X_1$, $X_2$, $Y_1$, and $Y_2$ be $n$-length sequences of independent random variables satisfying Assumption 1. Assume $\psi : \mathbb{R} \to \mathbb{R}$ is $C^4$ with $\|\psi^{(m)}\| \leq C$. Then

$$|E_X - E_Y| \leq \frac{2C}{3} \cdot 9^k \cdot \sum_{t=1}^{n} (\text{Inf}_t[f]^2 + \text{Inf}_t[g]^2 + 2 \text{Inf}_t[h]^2),$$

where $E_X$ and $E_Y$ are as defined in Theorem 4.

Corollary 5 (Separable BVIP-2). Let $F$ be an $n$-bivariate multilinear polynomial which is separable into $f$, $g$, and $h$, each of which is of degree at most $k$. Let $X_1$, $X_2$, $Y_1$, and $Y_2$ be $n$-length sequences of independent random variables satisfying Assumption 1. Assume $\psi : \mathbb{R} \to \mathbb{R}$ is $C^4$ with $\|\psi^{(m)}\| \leq C$. Then

$$|E_X - E_Y| \leq \frac{C}{6} \cdot 9^k \cdot \sum_{t=1}^{n} (\text{Inf}_t[f]^2 + \text{Inf}_t[g]^2 + 2 \text{Inf}_t[h]^2),$$

where $E_X$ and $E_Y$ are as defined in Theorem 4.

Clearly, for separable functions BVIP-1 yields a bound which is asymptotically tighter than that of BVIP-2 by a factor of $9^{k}$. This is due to the fact that $|T_2(S_1)|$ and $|T_1(S_2)|$ are constants for the case of separable functions. Note that this is not a general phenomenon: we can define functions such that $|T_2(S_1)|, |T_1(S_2)| \geq 9^k$, in which case BVIP-2 would provide a tighter bound. Nonetheless, for bivariate functions in which the interaction between inputs is not too strong or for functions of high degree, BVIP-1 will be tighter than the naive baseline of BVIP-2.

The fact that BVIP-1 is looser than BVIP-2 for some functions is evidence that our analysis is not perfect. It is left to future work to investigate and quantify the effect of the maximum degree and the interaction of the two inputs on the relative performance of these invariance principles. Furthermore, it is possible that other methods for proving the BIP would naturally lead to other bivariate invariance principles which may further elucidate this tradeoff or reveal new aspects of the problem. Finally, we also note that the bivariate method in this paper could potentially be extended to address multivariate, multilinear polynomials.
REFERENCES

[1] S. Skyum and L. G. Valiant, “A complexity theory based on Boolean algebra,” Journal of the ACM, vol. 32, no. 2, pp. 484–502, 1985.

[2] L. Babai, P. Frankl, and J. Simon, “Complexity classes in communication complexity theory,” in 27th Annual Symposium on Foundations of Computer Science (SFCS 1986), 1986, pp. 337–347.

[3] C. Carlet, Boolean Functions for Cryptography and Coding Theory. Cambridge University Press, 2021.

[4] T. W. Cusick and P. Stanica, Cryptographic Boolean Functions and Applications. Academic Press, 2017.

[5] G. Kalai and E. Mossel, “Sharp thresholds for monotone non-Boolean functions and social choice theory,” Mathematics of Operations Research, vol. 40, no. 4, pp. 915–925, 2015.

[6] A. M. Alturki and A. M. Ali Rushdi, “Weighted voting systems: A threshold-Boolean perspective,” Journal of Engineering Research, vol. 4, no. 1, pp. 1–19, 2016.

[7] C. E. Shannon, “A symbolic analysis of relay and switching circuits,” Electrical Engineering, vol. 57, no. 12, pp. 713–723, 1938.

[8] R. E. Bryant, “On the complexity of VLSI implementations and graph representations of Boolean functions with application to integer multiplication,” IEEE transactions on Computers, vol. 40, no. 2, pp. 205–213, 1991.

[9] E. Mossel, R. O’Donnell, and K. Oleszkiewicz, “Noise stability of functions with low influences: invariance and optimality,” in 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS’05), 2005, pp. 21–30.

[10] A. C. Berry, “The accuracy of the Gaussian approximation to the sum of independent variates,” Transactions of the american mathematical society, vol. 49, no. 1, pp. 122–136, 1941.

[11] C.-G. Esseen, “On the Liapounoff limit of error in the theory of probability,” Arkiv för matematik, astronomi och fysik, vol. 28, no. 9, pp. 1–19, 1942.

[12] L. S. Penrose, “The elementary statistics of majority voting,” Journal of the Royal Statistical Society, vol. 109, no. 1, pp. 53–57, 1946.

[13] V. Abdurashitov, M. Mérard, and D. Moshkovitz, “Matched filter decoding of random binary and gaussian codes in broadband gaussian channel,” in 2013 IEEE International Symposium on Information Theory. IEEE, 2013, pp. 2528–2532.

[14] R. G. L. D’Oliveira, S. El Rouayheb, and M. Mérard, “The computational wiretap channel,” in 2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2018, pp. 1136–1140.

[15] T. M. Cover and J. A. Thomas, Elements of Information Theory. John Wiley & Sons, 2006.

[16] R. O’Donnell, Analysis of Boolean Functions. Cambridge University Press, 2014.