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
Email: {amariona, homaesf, rafaeld, medard}@mit.edu

***Abstract*—The Basic Invariance Principle (BIP) is a nonlinear generalization of the Berry-Esseen Theorem which bounds the error in approximating the value of a multilinear polynomial over an n-length sequence of random variables by the same polynomial over a differently distributed n-length sequence. We present a generalization of the Basic Invariance Principle for bivariate multilinear polynomials, i.e. polynomials over two n-length sequences of random variables. This bivariate invariance principle is an iterative application of the BIP, using it to bound in turn the error from replacing each of the two input sequences. In order to prove this invariance principle, we derive a version of the BIP for random multilinear polynomials, i.e. polynomials whose coefficents are random variables.**

As a benchmark, we also state a naive bivariate 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**

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]. In this work, we primarily treat real-valued Boolean functions, which take the form $f : \{-1, 1\}^n \to \mathbb{R}$.

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. Given a real-valued Boolean function $f : \{-1, 1\}^n \to \mathbb{R}$ and two differently distributed sequences of independent random variables $\mathbf{X} = (X_1, \dots, X_n)$ and $\mathbf{Y} = (Y_1, \dots, Y_n)$, the BIP provides an explicit bound on the expected difference between $f(\mathbf{X})$ and $f(\mathbf{Y})$. This difference can be interpreted as the expected error incurred by approximating $f(\mathbf{X})$ as $f(\mathbf{Y})$.

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. 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 consequence of the BIP is the ability to "replace bits by Gaussians:" whether the input is a sequence of uniform random bits or a sequence of standard Gaussians[1], the expected output of a low-influence function does not change too much.

The BIP treats functions of a single sequence of random variables. One natural generalization of the BIP would be an invariance principle which treats functions of two sequences of random variables. 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 which directly from the BIP 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 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 $F(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) + g(\mathbf{y}) + h(\mathbf{xy})$, which we term *separable* functions. These functions of particular interesting because they generalize many different notions of noise that arise in communication channels.

The remainder of this paper is organized as follows. In Section II, we summarize key concepts and results from analysis of Boolean functions and formalize our notation for bivariate functions. In Section III, we define multilinear random polynomials and propose a version of the BIP that can be applied to those random functions in anticipation of Section IV, in which we state and prove BVIP-1. In Section V, we compare the bounds offered by the two principles, present corollary invariance principles for the special case of separable functions, and offer some 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.

[1] We gloss over here the technicalities of using real-valued sequences as the input to a function $f : \{-1, 1\}^n \to \mathbb{R}$. See Section II for more detail.

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 can specify the coordinates of a vector like $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. 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 we denote its power set with $2^{[n]}$.

A. Results from Analysis of Boolean Functions

First and foremost, we will be dealing with Boolean functions throughout this work in terms of their multilinear polynomial expansions.

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

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

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 $\widehat{f} : 2^{[n]} \to \mathbb{R}$. 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

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

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 $\mathbf{x}$ if negating x_i negates the output $f(\mathbf{x})$.

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

$$\mathbf{Inf}_i[f] = \Pr_{\mathbf{x} \sim \{-1,1\}^n} \big[f(\mathbf{x}) \neq f(\mathbf{x}^{\oplus i}) \big],$$

where

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

Informally, if we consider f to be a voting rule in a two-party election, the influence of the ith coordinate can be thought of as the "influence" or "power" of the ith 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 $f : \{-1, 1\}^n \to \mathbb{R}$ and $i \in [n]$, the influence of coordinate i on f is

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

This definition is given for Boolean functions, but we will also be interested in the influence of general multilinear polynomials whose inputs are arbitrary real-valued sequences. We will also use Definition 2 for such functions, a choice which is justified by Lemma 2 below.[2]

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.[3]

Assumption 1. The random variable X_i satisfies $\mathbf{E}[X_i] = 0$, $\mathbf{E}\big[X_i^2\big] = 1$, $\mathbf{E}\big[X_i^3\big] = 0$, and $\mathbf{E}\big[X_i^4\big] \leq 9$.

Two examples of random variables satisfying Assumption 1 are a uniform ± 1 random bit and a standard Gaussian.

The following two lemmas are used to prove the BIP using the replacement method, and we will use them to similar effect in Section III. Lemma 1 is one of the simpler hypercontractivity results.

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

$$\mathbf{E}\big[f(\mathbf{X})^4\big] \leq 9^k \cdot \mathbf{E}\big[f(\mathbf{X})^2\big]^2.$$

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

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

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

$$\mathbf{E}\big[f(\mathbf{X})^2\big] = \sum_{S \subseteq [n]} \widehat{f}(S)^2.$$

This leads us to the formal statement of the BIP.

Theorem 2 (BIP). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be an n-variate multilinear polynomial of degree at most k. Let $\mathbf{X}$ and $\mathbf{Y}$ be n-length sequences of random variables satisfying Assumption 1. Let $\psi : \mathbb{R} \to \mathbb{R}$ be $\mathcal{C}^4$, i.e. the derivatives $\psi', \ldots, \psi''''$ exist and are continuous, with $\|\psi''''\|_\infty \leq C$. Then*

$$|\mathbf{E}[\psi(f(\mathbf{X}))] - \mathbf{E}[\psi(f(\mathbf{Y}))]| \leq \frac{C}{12} \cdot 9^k \cdot \sum_{t=1}^{n} \mathbf{Inf}_t[f]^2.$$

[2] Lemma 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 [13, ch. 8.2] for more detail.

[3] In fact, a slightly different form of the BIP does hold 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 ψ 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 $\mathbf{E}[\psi(X)] \approx \mathbf{E}[\psi(Y)]$ for $\psi(s) = 1_{s \leq u}$ for all $u \in \mathbb{R}$. The BIP is powerful enough to give bounds on cdf-closeness and many other notions of closeness.[4]

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 $\mathbf{x}_1 = (x_{1,1}, \ldots, x_{1,n})$ and $\mathbf{x}_2 = (x_{2,1}, \ldots, x_{2,n})$ is a function of the form

$$f(\mathbf{x}_1, \mathbf{x}_2) = \sum_{S_1, S_2 \subseteq [n]} \widehat{f}(S_1, S_2) \prod_{i \in S_1} x_{1,i} \prod_{j \in S_2} x_{2,j}.$$

The form given in Definition 3 suggests that $\widehat{f}(S_1, S_2)$ is Fourier coefficient. Indeed, if we consider f to instead be a function of the concatenated sequence $\mathbf{x} = \mathbf{x}_1 \| \mathbf{x}_2$, then $\widehat{f}(S_1, S_2)$ 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 subleties 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 $f_{\mathbf{Z}} : \mathbb{R}^n \to \mathbb{R}$ is a multilinear polynomial over the sequence of indeterminates $\mathbf{x} = (x_1, \ldots, x_n)$ whose coefficients $\widehat{f_{\mathbf{Z}}}(S) \in \mathbb{R}$ are random variables which are wholly determined by the random variable $\mathbf{Z}$:

$$f_{\mathbf{Z}}(\mathbf{x}) = \sum_{S \subseteq [n]} \widehat{f_{\mathbf{Z}}}(S) \prod_{i \in S} x_i.$$

The influence of coordinate i on $f_{\mathbf{Z}}$ is a random variable which is defined to be

$$\mathbf{Inf}_i[f_{\mathbf{Z}}] = \sum_{S \ni i} \widehat{f_{\mathbf{Z}}}(S)^2.$$

We think of $\mathbf{Z}$ as the random variable which controls $f_{\mathbf{Z}}$ or, alternatively, which describes the randomness of $f_{\mathbf{Z}}$. We will sometimes refer to random multilinear polynomials simply as random functions; such functions will always be univariate.

[4]Note that continuity of ψ'''' is required. Smoothing techniques can be used to approximate functions like $\psi(s) = 1_{s \leq u}$. There is of course a tradeoff between the quality of the approximation and the magnitude of the fourth derivative of the smoothed function. See [13, ch. 11] for more detail.

Theorem 3 is a version of the BIP for random functions. Its proof closely follows that of the BIP, except that we need the appropriate corollaries of Lemma 1 and Lemma 2 for random functions.

Corollary 1. *Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sequence of independent but not necessarily identically distributed random variables satisfying the requirement that $\mathbf{E}[X_i] = \mathbf{E}\big[X_i^3\big] = 0$ and $\mathbf{E}\big[X_i^4\big] \leq 9\,\mathbf{E}\big[X_i^2\big]^2$. Let $f_{\mathbf{Z}}$ be a random multilinear polynomial of degree at most k. Then*

$$\mathop{\mathbf{E}}_{\mathbf{X},\mathbf{Z}}\big[f_{\mathbf{Z}}(\mathbf{X})^4\big] \leq 9^k \cdot \mathop{\mathbf{E}}_{\mathbf{X},\mathbf{Z}}\big[f_{\mathbf{Z}}(\mathbf{X})^2\big]^2.$$

Proof. We expand the expectation over $\mathbf{Z}$ using the law of total expectation. Without loss of generality, assume that $\mathbf{Z}$ is a random variable over a discrete sample space $\mathcal{Z}$. Then we can write

$$\begin{aligned}
\mathop{\mathbf{E}}_{\mathbf{X},\mathbf{Z}}\big[f_{\mathbf{Z}}(\mathbf{X})^4\big] &= \sum_{\mathbf{z} \in \mathcal{Z}} \Pr\{\mathbf{Z} = \mathbf{z}\} \mathop{\mathbf{E}}_{\mathbf{X}}\big[f_{\mathbf{Z}}(\mathbf{X})^4 \mid \mathbf{Z} = \mathbf{z}\big] \\
&= \sum_{\mathbf{z} \in \mathcal{Z}} \Pr\{\mathbf{Z} = \mathbf{z}\} \mathop{\mathbf{E}}_{\mathbf{X}}\big[f_{\mathbf{z}}(\mathbf{X})^4\big],
\end{aligned}$$

where $f_{\mathbf{z}}$ is the function $f_{\mathbf{Z}}$ given that $\mathbf{Z} = \mathbf{z}$. Conditioning on $\mathbf{Z} = \mathbf{z}$ fixes the coefficients of $f_{\mathbf{Z}}$, allowing us to apply Lemma 1 directly.

$$\begin{aligned}
\mathop{\mathbf{E}}_{\mathbf{X},\mathbf{Z}}\big[f_{\mathbf{Z}}(\mathbf{X})^4\big] &\leq \sum_{\mathbf{z} \in \mathcal{Z}} \Pr\{\mathbf{Z} = \mathbf{z}\} \cdot 9^k \cdot \mathop{\mathbf{E}}_{\mathbf{X}}\big[f_{\mathbf{z}}(\mathbf{X})^2\big]^2 \\
&= 9^k \cdot \mathop{\mathbf{E}}_{\mathbf{X},\mathbf{Z}}\big[f_{\mathbf{Z}}(\mathbf{X})^2\big]^2.
\end{aligned}$$

If $\mathcal{Z}$ is a continuous space, then the sums over $\mathcal{Z}$ will be replaced by integrals, by the argument is otherwise the same.□

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

$$\mathop{\mathbf{E}}_{\mathbf{X},\mathbf{Z}}\big[f_{\mathbf{Z}}(\mathbf{X})^2\big] = \mathop{\mathbf{E}}_{\mathbf{Z}}\left[\sum_{S \subseteq [n]} \widehat{f_{\mathbf{Z}}}(S)^2\right].$$

Proof. As in the proof of Corollary 1, we expand the expectation over $\mathbf{Z}$ using the law of total expectation.

$$\begin{aligned}
\mathop{\mathbf{E}}_{\mathbf{X},\mathbf{Z}}\big[f_{\mathbf{Z}}(\mathbf{X})^2\big] &= \sum_{\mathbf{z} \in \mathcal{Z}} \Pr\{\mathbf{Z} = \mathbf{z}\} \mathop{\mathbf{E}}_{\mathbf{X}}\big[f_{\mathbf{Z}}(\mathbf{X})^2 \mid \mathbf{Z} = \mathbf{z}\big] \\
&= \sum_{\mathbf{z} \in \mathcal{Z}} \Pr\{\mathbf{Z} = \mathbf{z}\} \mathop{\mathbf{E}}_{\mathbf{X}}\big[f_{\mathbf{z}}(\mathbf{X})^2\big].
\end{aligned}$$

Conditioning on $\mathbf{Z} = \mathbf{z}$, we apply Lemma 2 directly.

$$\begin{aligned}
\mathop{\mathbf{E}}_{\mathbf{X},\mathbf{Z}}\big[f_{\mathbf{Z}}(\mathbf{X})^2\big] &= \sum_{\mathbf{z} \in \mathcal{Z}} \left(\Pr\{\mathbf{Z} = \mathbf{z}\} \sum_{S \subseteq [n]} \widehat{f_{\mathbf{z}}}(S)^2\right) \\
&= \mathop{\mathbf{E}}_{\mathbf{Z}}\left[\sum_{S \subseteq [n]} \widehat{f_{\mathbf{Z}}}(S)^2\right].
\end{aligned}$$

As before, if $\mathcal{Z}$ is a continuous space, we can replace the sums of $\mathcal{Z}$ with integrals without issue. $\square$

Theorem 3 (BIP for Random Functions). *Let $f_{\mathbf{Z}}$ be a random n-variate multilinear polynomial of degree at most k. Let $\mathbf{X}$ and $\mathbf{Y}$ be n-length sequences of random variables satisfying Assumption 1. Let $\psi : \mathbb{R} \to \mathbb{R}$ be $\mathcal{C}^4$, i.e. the derivatives $\psi', \ldots, \psi''''$ exist and are continuous, with $\|\psi''''\|_\infty \leq C$. Then*

$$|\mathbf{E}[\psi(f_{\mathbf{Z}}(\mathbf{X}))] - \mathbf{E}[\psi(f_{\mathbf{Z}}(\mathbf{Y}))]| \leq \frac{C}{12} \cdot 9^k \cdot \sum_{t=1}^{n} \mathop{\mathbf{E}}_{\mathbf{Z}}[\mathbf{Inf}_t[f_{\mathbf{Z}}]]^2.$$

Proof. The proof closely follows the proof of the BIP given in [13, ch. 11.6], so we omit some exposition which can be found there. Nonetheless, for completeness we summarize the arguments and highlight the points where the random functions affect the process.

We use the replacement method and define

$$H_t = f_{\mathbf{z}}(Y_1, \ldots, Y_t, X_{t+1}, , \ldots, X_n),$$

such that $H_0 = f_{\mathbf{z}}(\mathbf{X})$ and $H_n = f_{\mathbf{z}}(\mathbf{Y})$. We show that

$$\left| \mathop{\mathbf{E}}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}[\psi(H_{t-1}) - \psi(H_t)] \right| \leq \frac{C}{12} \cdot 9^k \cdot \mathop{\mathbf{E}}_{\mathbf{Z}}[\mathbf{Inf}_t[f_{\mathbf{Z}}]]^2. \quad (1)$$

Summing over t and applying the triangle inequality will complete the proof.

Let the random functions $\mathrm{E}_t f_{\mathbf{Z}}$ and $\mathrm{D}_t f_{\mathbf{Z}}$ be defined as

$$\mathrm{E}_t f_{\mathbf{Z}}(x) = \sum_{S \not\ni t} \widehat{f_{\mathbf{Z}}}(S) \prod_{i \in S} x_i$$
$$\mathrm{D}_t f_{\mathbf{Z}}(x) = \sum_{S \ni t} \widehat{f_{\mathbf{Z}}}(S) \prod_{i \in S \setminus \{t\}} x_i,$$

such that $f_{\mathbf{z}}(\mathbf{x}) = \mathrm{E}_t f_{\mathbf{Z}}(\mathbf{x}) + x_t \mathrm{D}_t f_{\mathbf{Z}}(\mathbf{x})$. Since neither $\mathrm{E}_t f_{\mathbf{Z}}$ nor $\mathrm{D}_t f_{\mathbf{Z}}$ depends on x_t, we can define

$$U_t = \mathrm{E}_t f_{\mathbf{Z}}(Y_1, \ldots, Y_{t-1}, \cdot, X_{t+1}, \ldots, X_n),$$
$$\Delta_t = \mathrm{D}_t f_{\mathbf{Z}}(Y_1, \ldots, Y_{t-1}, \cdot, X_{t+1}, \ldots, X_n),$$

so that

$$H_{t-1} = U_t + \Delta_t X_t, \quad H_t = U_t + \Delta_t Y_t.$$

We can then bound (1) by taking 3rd-order Taylor expansions of $\psi(H_{t-1})$ and $\psi(H_t)$ and then taking the difference between them. After subtracting and taking expectations over $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, the 0th-order terms cancel directly, and the 1st-, 2nd-, and 3rd-order terms cancel because X_t and Y_t are independent of U_t and Δ_t and X_t and Y_t have matching 1st, 2nd, and 3rd moments. For the 4th-order error term, we apply the triangle inequality and make use of the assumption that $|\psi''''(U_t^*)|, |\psi''''(U_t^{**})| \leq C$ to upper bound the left-hand side of (1) by

$$\frac{C}{24} \cdot \left(\mathop{\mathbf{E}}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}\left[(\Delta_t X_t)^4\right] + \mathop{\mathbf{E}}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}\left[(\Delta_t Y_t)^4\right] \right).$$

All that remains is to bound

$$\mathop{\mathbf{E}}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}\left[(\Delta_t X_t)^4\right], \mathop{\mathbf{E}}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}\left[(\Delta_t Y_t)^4\right] \leq 9^k \cdot \mathop{\mathbf{E}}_{\mathbf{Z}}[\mathbf{Inf}_t[f_{\mathbf{Z}}]]^2,$$

which can be done using Corollary 1. We give details for the case of $\mathbf{E}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}\left[(\Delta_t X_t)^4\right]$; the case for $\mathbf{E}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}\left[(\Delta_t Y_t)^4\right]$ is identical. Define

$$\mathrm{L}_t f_{\mathbf{Z}}(\mathbf{x}) = x_t \mathrm{D}_t f_{\mathbf{Z}}(\mathbf{x}) = \sum_{S \ni t} \widehat{f}(S) \prod_{i \in S} x_i.$$

Then, $\Delta_t X_t = \mathrm{L}_t f_{\mathbf{Z}}(Y_1, \ldots, X_t, \ldots, X_n)$. Since $\mathrm{L}_t f_{\mathbf{Z}}$ has degree at most k we can apply Corollary 1 to obtain

$$\mathop{\mathbf{E}}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}\left[(\Delta_t X_t)^4\right] \leq 9^k \cdot \mathop{\mathbf{E}}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}\left[(\Delta_t X_t)^2\right]^2. \quad (2)$$

Finally, since the elements of $\mathbf{X}$ and $\mathbf{Y}$ all have mean 0 and 2nd moment 1, by Corollary 2

$$\begin{aligned} \mathop{\mathbf{E}}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}\left[(\Delta_t X_t)^2\right] &= \mathop{\mathbf{E}}_{\mathbf{Z}}\left[\sum_{S \subseteq [n]} \widehat{\mathrm{L}_t f_{\mathbf{Z}}}(S)^2\right] \\ &= \mathop{\mathbf{E}}_{\mathbf{Z}}\left[\sum_{S \ni t} \widehat{f_{\mathbf{Z}}}(S)^2\right] \\ &= \mathop{\mathbf{E}}_{\mathbf{Z}}[\mathbf{Inf}_t[f_{\mathbf{Z}}]]. \end{aligned} \quad (3)$$

Combining (2) and (3), we have that

$$\mathop{\mathbf{E}}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}\left[(\Delta_t X_t)^4\right] \leq 9^k \cdot \mathop{\mathbf{E}}_{\mathbf{Z}}[\mathbf{Inf}_t[f_{\mathbf{Z}}]]^2,$$

which completes the proof. $\square$

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:*

$$f(\mathbf{x}_1, \mathbf{x}_2) = \sum_{S_1, S_2 \subseteq [n]} \widehat{f}(S_1, S_2) \prod_{i \in S_1} x_{1,i} \prod_{j \in S_2} x_{2,j},$$

where $\widehat{f}(S_1, S_2) = 0$ if $|S_1| > k$ or $|S_2| > k$. Let $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{Y}_1$, and $\mathbf{Y}_2$ be n-length sequences of independent random variables satisfying Assumption 1. Assume $\psi : \mathbb{R} \to \mathbb{R}$ is $\mathcal{C}^4$ with $\|\psi''''\|_\infty \leq C$. Then

$$|E_X - E_Y| \leq \frac{C}{12} \cdot 9^k \cdot \sum_{t=1}^{n} \left(\widetilde{\Sigma}_{1,t}^2 + \widetilde{\Sigma}_{2,t}^2 \right), \quad (4)$$

where

$$\begin{aligned} E_X &= \mathop{\mathbf{E}}_{\mathbf{X}_1,\mathbf{X}_2}[\psi(f(\mathbf{X}_1, \mathbf{X}_2))] \\ E_Y &= \mathop{\mathbf{E}}_{\mathbf{Y}_1,\mathbf{Y}_2}[\psi(f(\mathbf{Y}_1, \mathbf{Y}_2))] \\ \widetilde{\Sigma}_{1,t} &= \sum_{S_1 \ni t} |T_2(S_1)| \sum_{S_2 \in T_2(S_1)} \widehat{f}(S_1, S_2)^2 \\ \widetilde{\Sigma}_{2,t} &= \sum_{S_2 \ni t} |T_1(S_2)| \sum_{S_1 \in T_1(S_2)} \widehat{f}(S_1, S_2)^2, \end{aligned}$$

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

$$\begin{aligned} T_2(S_1) &= \left\{ S_2 \subseteq [n] : |S_2| \le k,\ \widehat{f}(S_1,S_2) \neq 0 \right\} \\ T_1(S_2) &= \left\{ S_1 \subseteq [n] : |S_1| \le k,\ \widehat{f}(S_1,S_2) \neq 0 \right\}. \end{aligned}$$

Proof. As described above, our strategy is to define random functions $f_{\mathbf{X}_2}$ and $f_{\mathbf{Y}_1}$ such that $f_{\mathbf{X}_2}(\mathbf{X}_1) = f(\mathbf{X}_1, \mathbf{X}_2)$ and $f_{\mathbf{Y}_1}(\mathbf{X}_2) = f(\mathbf{Y}_1, \mathbf{X}_2)$. Applying Theorem 3 to $f_{\mathbf{X}_2}$ bounds the error incurred by replacing $\mathbf{X}_1$ with $\mathbf{Y}_1$; an application of the same theorem to $f_{\mathbf{Y}_1}$ bounds the error in replacing $\mathbf{X}_2$ and $\mathbf{Y}_2$. 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 $f_{\mathbf{X}_2} : \mathbb{R}^n \to \mathbb{R}$ be defined as

$$\begin{aligned} f_{\mathbf{X}_2}(\mathbf{t}) &= f(\mathbf{t}, \mathbf{X}_2) \\ &= \sum_{S_1 \subseteq [n]} \left[\sum_{S_2 \subseteq [n]} \widehat{f}(S_1,S_2) \prod_{j \in S_2} X_{2,j} \right] \prod_{i \in S_1} t_i \\ &= \sum_{S_1 \subseteq [n]} \widehat{f_{\mathbf{X}_2}}(S_1) \prod_{i \in S_1} t_i. \end{aligned}$$

Similarly, let $f_{\mathbf{Y}_1}$ be defined as

$$\begin{aligned} f_{\mathbf{Y}_1}(\mathbf{t}) &= f(\mathbf{Y}_1, \mathbf{t}) \\ &= \sum_{S_2 \subseteq [n]} \left[\sum_{S_1 \subseteq [n]} \widehat{f}(S_1,S_2) \prod_{i \in S_1} Y_{1,j} \right] \prod_{i \in S_2} t_i \\ &= \sum_{S_2 \subseteq [n]} \widehat{f_{\mathbf{Y}_1}}(S_2) \prod_{i \in S_2} t_i. \end{aligned}$$

Note that both $f_{\mathbf{X}_2}$ and $f_{\mathbf{Y}_1}$ are of degree at most k and that $f_{\mathbf{X}_2}(\mathbf{Y}_1) = f_{\mathbf{Y}_1}(\mathbf{X}_2)$. From the definitions of E_X and E_Y,

$$\begin{aligned} E_X &= \mathop{\mathbf{E}}_{\mathbf{X}_1,\mathbf{X}_2}[\psi(f_{\mathbf{X}_2}(\mathbf{X}_1))] \\ E_Y &= \mathop{\mathbf{E}}_{\mathbf{Y}_1,\mathbf{Y}_2}[\psi(f_{\mathbf{Y}_1}(\mathbf{Y}_2))]. \end{aligned}$$

By analogy, let

$$\begin{aligned} E_{XY} &= \mathop{\mathbf{E}}_{\mathbf{Y}_1,\mathbf{X}_2}[\psi(f(\mathbf{Y}_1,\mathbf{X}_2)] \\ &= \mathop{\mathbf{E}}_{\mathbf{Y}_1,\mathbf{X}_2}[\psi(f_{\mathbf{X}_2}(\mathbf{Y}_1)] \\ &= \mathop{\mathbf{E}}_{\mathbf{Y}_1,\mathbf{X}_2}[\psi(f_{\mathbf{Y}_1}(\mathbf{X}_2)]. \end{aligned}$$

We upper bound the quantity of interest as

$$|E_X - E_Y| \le |E_X - E_{XY}| + |E_{XY} - E_Y|. \tag{5}$$

Applying Theorem 3 to each term on the right-hand side of (5) yields

$$\begin{aligned} |E_X - E_{XY}| &= \left| \mathop{\mathbf{E}}_{\mathbf{X}_1,\mathbf{X}_2}[\psi(f_{\mathbf{X}_2}(\mathbf{X}_1))] - \mathop{\mathbf{E}}_{\mathbf{Y}_1,\mathbf{X}_2}[\psi(f_{\mathbf{X}_2}(\mathbf{Y}_1)] \right| \\ &\le \frac{C}{12} \cdot 9^k \cdot \sum_{t=1}^{n} \mathop{\mathbf{E}}_{\mathbf{X}_2}[\mathbf{Inf}_t[f_{\mathbf{X}_2}]]^2 && (6) \\ |E_{XY} - E_Y| &= \left| \mathop{\mathbf{E}}_{\mathbf{Y}_1,\mathbf{X}_2}[\psi(f_{\mathbf{Y}_1}(\mathbf{X}_2))] - \mathop{\mathbf{E}}_{\mathbf{Y}_1,\mathbf{Y}_2}[\psi(f_{\mathbf{Y}_1}(\mathbf{Y}_2)] \right| \\ &\le \frac{C}{12} \cdot 9^k \cdot \sum_{t=1}^{n} \mathop{\mathbf{E}}_{\mathbf{Y}_1}[\mathbf{Inf}_t[f_{\mathbf{Y}_1}]]^2. && (7) \end{aligned}$$

All that remains is to bound the expected influences of $f_{\mathbf{X}_2}$ and $f_{\mathbf{Y}_1}$. We handle the case of $f_{\mathbf{X}_2}$ explicitly, with the argument for $f_{\mathbf{Y}_1}$ being identical. For convenience, define

$$\sigma_2(S_1) = \sum_{S_2 \in T_2(S_1)} \widehat{f}(S_1,S_2)^2.$$

We have

$$\begin{aligned} \mathop{\mathbf{E}}_{\mathbf{X}_2}[\mathbf{Inf}_t[f_{\mathbf{X}_2}]] &= \mathop{\mathbf{E}}_{\mathbf{X}_2}\left[\sum_{S_1 \ni t} \widehat{f_{\mathbf{X}_2}}(S_1)^2 \right] \\ &= \mathop{\mathbf{E}}_{\mathbf{X}_2}\left[\sum_{S_1 \ni t} \left(\sum_{S_2 \subseteq [n]} \widehat{f}(S_1,S_2) \prod_{j \in S_2} X_{2,j} \right)^2 \right] \\ &= \sum_{S_1 \ni t} \mathop{\mathbf{E}}_{\mathbf{X}_2}\left[\left(\sum_{S_2 \in T_2(S_1)} \widehat{f}(S_1,S_2) \prod_{j \in S_2} X_{2,j} \right)^2 \right] && (8) \\ &\le \sum_{S_1 \ni t} \mathop{\mathbf{E}}_{\mathbf{X}_2}\left[\sigma_2(S_1) \left(\sum_{S_2 \in T_2(S_1)} \prod_{j \in S_2} X_{2,j}^2 \right) \right] && (9) \\ &= \sum_{S_1 \ni t} \sigma_2(S_1) \left(\sum_{S_2 \in T_2(S_1)} \mathop{\mathbf{E}}_{\mathbf{X}_2}\left[\prod_{j \in S_2} X_{2,j}^2 \right] \right) && (10) \\ &= \sum_{S_1 \ni t} \sigma_2(S_1) \left(\sum_{S_2 \in T_2(S_1)} \prod_{j \in S_2} \mathop{\mathbf{E}}_{\mathbf{X}_2}[X_{2,j}^2] \right) && (11) \\ &= \sum_{S_1 \ni t} |T_2(S_1)| \cdot \sigma_2(S_1) && (12) \\ &= \widetilde{\Sigma}_{1,t}, && (13) \end{aligned}$$

where (8) follows from linearity of expectation and the fact that for a given S_1, we have $\widehat{f}(S_1,S_2) \neq 0$ only if $S_2 \in T_2(S_2)$; (9) follows from the Cauchy-Schwarz inequality; (10) again follows from linearity of expectation; (11) follows from the assumption that the elements of $\mathbf{X}_2$ are independent; and (12) follows from the assumption that $\mathbf{E}[X_{2,j}^2] = 1$ for all $j \in [n]$. The same argument applied to $f_{\mathbf{Y}_1}$ gives

$$\mathop{\mathbf{E}}_{\mathbf{Y}_1}[\mathbf{Inf}_t[f_{\mathbf{Y}_1}]] \le \widetilde{\Sigma}_{2,t}. \tag{14}$$

Substituting (13) and (14) into (6) and (7) respectively yields

$$|E_X - E_{XY}| \leq \frac{C}{12} \cdot 9^k \cdot \sum_{t=1}^{n} \widetilde{\Sigma}_{1,t}^2$$

$$|E_{XY} - E_Y| \leq \frac{C}{12} \cdot 9^k \cdot \sum_{t=1}^{n} \widetilde{\Sigma}_{2,t}^2,$$

which, combined with (5), completes the proof. □

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 would take as its input the concatenation of the two sequences which were 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:*

$$f(\mathbf{x}_1, \mathbf{x}_2) = \sum_{S_1, S_2 \subseteq [n]} \widehat{f}(S_1, S_2) \prod_{i \in S_1} x_{1,i} \prod_{j \in S_2} x_{2,j},$$

where $\widehat{f}(S_1, S_2) = 0$ if $|S_1| > k$ or $|S_2| > k$. Let $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{Y}_1$, and $\mathbf{Y}_2$ be n-length sequences of independent random variables satisfying Assumption 1. Assume $\psi : \mathbb{R} \to \mathbb{R}$ is $\mathcal{C}^4$ with $\|\psi''''\|_\infty \leq C$. Then

$$|E_X - E_Y| \leq \frac{C}{12} \cdot 9^{2k} \cdot \sum_{t=1}^{n} (\Sigma_{1,t}^2 + \Sigma_{2,t}^2), \tag{15}$$

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

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

$$\Sigma_{2,t} = \sum_{S_2 \ni t} \sum_{S_1 \in T_1(S_2)} \widehat{f}(S_1, S_2)^2.$$

Proof. The strategy is to define a univariate function g which is equivalent to f when the two input sequences are considered as a single sequence so that we may then apply the BIP to g. Given a particular subset $S \subseteq [2n]$, let $S_1^* = S \cap [n]$ and $S_2^* = S \cap \{n+1, \ldots, 2n\}$. Then let g be a $2n$-variate multilinear polynomial of degree such that

$$g(\mathbf{x}) = \sum_{S \subseteq [n]} \widehat{f}(S_1^*, S_2^*) \prod_{i \in S} x_i.$$

For any n-length sequences $\mathbf{x}_1$ and $\mathbf{x}_2$, it is clear that $g(\mathbf{x}_1 \| \mathbf{x}_2) = f(\mathbf{x}_1, \mathbf{x}_2)$, where $\mathbf{x}_1 \| \mathbf{x}_2$ is the concatentation of $\mathbf{x}_1$ and $\mathbf{x}_2$. Furthermore, since $\widehat{f}(S_1, S_2) = 0$ when $|S_1| > k$ or $|S_2| > k$, by construction g is at most $2k$.

Applying the BIP to g for the concatenations $\mathbf{X} = \mathbf{X}_1 \| \mathbf{X}_2$ and $\mathbf{Y} = \mathbf{Y}_1 \| \mathbf{Y}_2$

$$|\mathbf{E}[\psi(g(\mathbf{X}))] - \mathbf{E}[\psi(g(\mathbf{Y}))]| \leq \frac{C}{12} \cdot 9^{2k} \cdot \sum_{t=1}^{2n} \mathbf{Inf}_t[g]^2. \tag{16}$$

We now compute $\mathbf{Inf}_t[g]$ in terms of the coefficients of f. By definition, $\widehat{g}(S) = \widehat{f}(S_1^*, S_2^*)$. Thus, for $t \in [n]$,

$$\begin{aligned}
\mathbf{Inf}_t[g] &= \sum_{S \ni t} \widehat{G}(S)^2 \\
&= \sum_{S_1 \ni t} \sum_{S_2 \subseteq [n]} \widehat{f}(S_1, S_2)^2 \\
&= \sum_{S_1 \ni t} \sum_{S_2 \in T_2(S_1)} \widehat{f}(S_1, S_2)^2 \qquad (17) \\
&= \Sigma_{t,1}.
\end{aligned}$$

where (17) follows from the definition of the set $T_2(S_1)$. By a paralell argument, for $t \in \{n+12, \ldots, n\}$,

$$\begin{aligned}
\mathbf{Inf}_t[g] &= \sum_{S \ni t} \widehat{G}(S)^2 \\
&= \sum_{S_2 \ni t-n} \sum_{S_1 \subseteq [n]} \widehat{f}(S_1, S_2)^2 \\
&= \sum_{S_2 \ni t-n} \sum_{S_1 \in T_1(S_2)} \widehat{f}(S_1, S_2)^2 \qquad (18) \\
&= \Sigma_{t-n,2}.
\end{aligned}$$

Combining (17) and (18),

$$\begin{aligned}
\sum_{t=1}^{2n} \mathbf{Inf}_t[g]^2 &= \sum_{t=1}^{n} \mathbf{Inf}_t[g]^2 + \sum_{t=n+1}^{2n} \mathbf{Inf}_t[g]^2 \\
&= \sum_{t=1}^{n} \Sigma_{t,1}^2 + \sum_{t=n+1}^{2n} \Sigma_{t-n,2}^2 \\
&= \sum_{t=1}^{n} \Sigma_{t,1}^2 + \Sigma_{t,2}^2. \qquad (19)
\end{aligned}$$

Substituting (19) into (16) yields the desired inequality after replacing $g(\mathbf{X})$ and $g(\mathbf{Y})$ on the left-hand side of (16) with $f(\mathbf{X}_1, \mathbf{X}_2)$ and $f(\mathbf{Y}_1, \mathbf{Y}_2)$. □

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 (4) and (15) respectively, the main differences are a factor of 9^k versus 9^{2k} and the quantity $\widetilde{\Sigma}_{i,t}$ versus $\Sigma_{i,t}$ (for $i \in \{1, 2\}$), which we recall are defined as (for $i = 1$)

$$\widetilde{\Sigma}_{1,t} = \sum_{S_1 \ni t} |T_2(S_1)| \sum_{S_2 \in T_2(S_1)} \widehat{f}(S_1, S_2)^2$$

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

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 $|(T_1(S_2)|$ 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 $\mathbf{X}_1$ and $\mathbf{X}_2$. If those cardinalities are large, then there are many terms of f in which some coordinates of $\mathbf{X}_2$ are multiplied with the coordinates of $\mathbf{X}_1$.

The question of whether BVIP-1 or BVIP-2 is tighter for a particular f is then a question of whether the inputs $\mathbf{X}_1$ and $\mathbf{X}_2$ interact enough to outweigh the extra factor of 9^k. As a concrete example of a family of functions for which BVIP-1 is always tighter than BVIP-2, we can 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

$$F(\mathbf{x}_1, \mathbf{x}_2) = f(\mathbf{x}_1) + g(\mathbf{x}_2) + h(\mathbf{x}_1\mathbf{x}_2).$$

For separable functions, the bounds of both bivariate invariance principles can be very 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. We states these bounds in the following two corollaries.

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 $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{Y}_1$, and $\mathbf{Y}_2$ be n-length sequences of independent random variables satisfying Assumption 1. Assume $\psi : \mathbb{R} \to \mathbb{R}$ is $\mathcal{C}^4$ with $\|\psi''''\| \le C$. Then*

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

where E_X and E_Y are as defined in Theorem 4.

Proof. By assumption, F is of the form

$$F(\mathbf{x}_1, \mathbf{x}_2) = f(\mathbf{x}_1) + g(\mathbf{x}_2) + h(\mathbf{x}_1\mathbf{x}_2).$$

Since f, g, and h are each of degree at most k, each term of F includes at most k elements from each input sequence. Thus, we can apply Theorem 4.

For separable functions, we can compute $T_2(S_1)$ and $T_1(S_2)$ directly. We have

$$T_2(S_1) = \{\emptyset, S_1\}, \quad T_1(S_2) = \{\emptyset, S_2\},$$

Hence, for a given $S \subseteq [n]$, the only (possibly) non-zero coefficients of F are

$$\widehat{F}(S, \emptyset) = \widehat{f}(S), \quad \widehat{F}(\emptyset, S) = \widehat{g}(S), \quad \widehat{F}(S, S) = \widehat{h}(S).$$

Computing $\widetilde{\Sigma}_{1,t}$, we have

$$\begin{aligned}
\widetilde{\Sigma}_{1,t} &= \sum_{S_1 \ni t} |T_2(S_1)| \sum_{S_2 \ni T_2(S_1)} \widehat{F}(S_1, S_2)^2 \\
&= \sum_{S_1 \ni t} 2\Big(\widehat{F}(S_1, \emptyset)^2 + \widehat{F}(S_1, S_1)^2\Big) \\
&= \sum_{S_1 \ni t} 2\Big(\widehat{f}(S_1)^2 + \widehat{h}(S_1)^2\Big) \\
&= 2\sum_{S_1 \ni t} \widehat{f}(S_1)^2 + 2\sum_{S_1 \ni t} \widehat{h}(S_1)^2 \\
&= 2\,\mathbf{Inf}_t[f] + 2\,\mathbf{Inf}_t[h].
\end{aligned}$$

Similarly, we have

$$\widetilde{\Sigma}_{2,t} = 2\,\mathbf{Inf}_t[g] + 2\,\mathbf{Inf}_t[h].$$

A simple application of Cauchy-Schwarz yields

$$\widetilde{\Sigma}_{1,t}^2 + \widetilde{\Sigma}_{2,t}^2 \le 8\,\mathbf{Inf}_t[f]^2 + 8\,\mathbf{Inf}_t[g]^2 + 16\,\mathbf{Inf}_t[h]^2. \quad (20)$$

Substituting (20) into the bound of Theorem 4 yields the desired inequality. □

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 $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{Y}_1$, and $\mathbf{Y}_2$ be n-length sequences of independent random variables satisfying Assumption 1. Assume $\psi : \mathbb{R} \to \mathbb{R}$ is $\mathcal{C}^4$ with $\|\psi''''\| \le C$. Then*

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

where E_X and E_Y are as defined in Theorem 4.

Proof. As in the proof of Corollary 4, we again have

$$T_2(S_1) = \{\emptyset, S_1\}, \quad T_1(S_2) = \{\emptyset, S_2\},$$

with the possibly non-zero coefficients for a given $S \subseteq [n]$ being

$$\widehat{F}(S, \emptyset) = \widehat{f}(S), \quad \widehat{F}(\emptyset, S) = \widehat{g}(S), \quad \widehat{F}(S, S) = \widehat{h}(S).$$

Computing $\Sigma_{1,t}$, we have

$$\begin{aligned}
\Sigma_{1,t} &= \sum_{S_1 \ni t} \sum_{S_2 \in T_2(S_1)} \widehat{F}(S_1, S_2)^2 \\
&= \sum_{S_1 \ni t} \widehat{F}(S_1, \emptyset)^2 + \widehat{F}(S_1, S_1)^2 \\
&= \sum_{S_1 \ni t} \widehat{f}(S_1)^2 + \sum_{S_1 \ni t} \widehat{h}(S_1)^2 \\
&= \mathbf{Inf}_t[f] + \mathbf{Inf}_t[h].
\end{aligned}$$

Similarly, we have

$$\Sigma_{2,t} = \mathbf{Inf}_t[g] + \mathbf{Inf}_t[h].$$

A simple application of Cauchy-Schwarz yields

$$\Sigma_{1,t}^2 + \Sigma_{2,t}^2 \le 2\,\mathbf{Inf}_t[f]^2 + 2\,\mathbf{Inf}_t[g]^2 + 4\,\mathbf{Inf}_t[h]^2. \quad (21)$$

Substituting (21) into the bound of Corollary 3 yields the desired inequality. □

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 $|T_2(S_1)|, |T_1(S_2)| \ge 9^k$, 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 BVIP2. A more quantitative characterization of this relationship, as well as generalizations of BVIP-1 to multivariate functions, is left to future work.

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] R. O'Donnell, *Analysis of Boolean Functions*. Cambridge University Press, 2014.