Cryptography and Communications (2022) 14:459–481
https://doi.org/10.1007/s12095-021-00529-4

Improving bounds on probabilistic affine tests to estimate the nonlinearity of Boolean functions

Ana Sălăgean[1] · **Pantelimon Stănică**[2]

Received: 16 November 2020 / Accepted: 4 August 2021 / Published online: 19 November 2021
© The Author(s) 2021

Abstract

In this paper we want to estimate the nonlinearity of Boolean functions, by probabilistic methods, when it is computationally very expensive, or perhaps not feasible to compute the full Walsh transform (which is the case for almost all functions in a larger number of variables, say more than 30). Firstly, we significantly improve upon the bounds of Zhang and Zheng (1999) on the probabilities of failure of affinity tests based on nonhomomorphicity, in particular, we prove a new lower bound that we have previously conjectured. This new lower bound generalizes the one of Bellare et al. (IEEE Trans. Inf. Theory **42**(6), 1781–1795 1996) to nonhomomorphicity tests of arbitrary order. Secondly, we prove bounds on the probability of failure of a proposed affinity test that uses the BLR linearity test. All these bounds are expressed in terms of the function's nonlinearity, and we exploit that to provide probabilistic methods for estimating the nonlinearity based upon these affinity tests. We analyze our estimates and conclude that they have reasonably good accuracy, particularly so when the nonlinearity is low.

Keywords Nonlinearity · Walsh transform · Probabilistic testing · Nonhomomorphicity

Mathematics Subject Classification (2010) 06E30 · 60C05 · 94A60 · 94D10

This article belongs to the Topical Collection: *Sequences and Their Applications III*
Guest Editors: Chunlei Li, Tor Helleseth and Zhengchun Zhou

This is a substantially revised and extended version of the article [11] that appeared in the proceedings of Sequences and Their Applications – SETA 2020. In particular, Section 4 and Proposition 13 are new.

✉ Ana Sălăgean
A.M.Salagean@lboro.ac.uk

Pantelimon Stănică
pstanica@nps.edu

[1] Department of Computer Science, Loughborough University, Loughborough, UK

[2] Applied Mathematics Department, Naval Postgraduate School, Monterey, CA 93943, USA

1 Introduction and motivation

Boolean functions are defined on a vector space over the binary finite field $\mathbb{F}_2$ with output in $\mathbb{F}_2$. For many cryptographic applications it is important that functions are not affine, and not even close (with respect to the Hamming distance, defined in (3)) to being affine. The nonlinearity of a function f, denoted $\mathrm{d}_{\mathcal{A}}(f)$, defined as the minimum Hamming distance to any affine function, is therefore an important cryptographic property. This indicator can be computed by using the Walsh transform (also called Walsh-Hadamard or discrete Fourier transform). The Walsh transform of a function f in n variables can be computed from its truth table by an algorithm similar to the fast Fourier transform in time $O(n2^n)$. Computing the Walsh transform is not feasible in practice when the number of variables is large (e.g., it is not feasible for functions in 80 variables; functions which model an output of a stream or block cipher as a function of the key would have a number of variables equal to the length of the key, i.e. at least 80 variables) and the function is given as a "black box" (or given by an algorithm or formula which is not amenable to simple manipulation for the purpose of computing the Walsh transform).

The motivation of this paper is to probabilistically estimate the nonlinearity of f to a reasonable degree of accuracy. The main idea is as follows. Consider a probabilistic test (we will see some examples shortly) which has a success/fail outcome based on the values of f at some fixed number k of points in $\mathbb{F}_2^n$ (f can therefore be given as a "black box" function). Denote by $T(f)$ the probability of failing the test (with the probability taken over all possible choices of k inputs in $\mathbb{F}_2^n$). We assume $T(f)$ is positively correlated, to some extent, with the nonlinearity $\mathrm{d}_{\mathcal{A}}(f)$, and can be bounded by some functions in $\mathrm{d}_{\mathcal{A}}(f)$, say $\mathrm{Lower}(\mathrm{d}_{\mathcal{A}}(f)) \leq T(f) \leq \mathrm{Upper}(\mathrm{d}_{\mathcal{A}}(f))$. If we can obtain $T(f)$ with reasonable accuracy by practical statistical testing (e.g. binomial proportion confidence interval), we can then estimate the nonlinearity as:

$$\mathrm{d}_{\mathcal{A}}(f) \in [\min(\mathrm{Upper}^{-1}(T(f))), \max(\mathrm{Lower}^{-1}(T(f)))], \tag{1}$$

(we use $F^{-1}(x)$ to denote the preimage of x under F), or, if the preimage has only one element,

$$\mathrm{d}_{\mathcal{A}}(f) \in [\mathrm{Upper}^{-1}(T(f)), \mathrm{Lower}^{-1}(T(f))]. \tag{2}$$

To obtain an accurate estimate, it is important that $T(f)$ depend strongly on $\mathrm{d}_{\mathcal{A}}(f)$ and that the bounds are very good. We will examine several probabilistic tests, improve some of the existing bounds, and analyze the accuracy of the resulting estimation.

The linearity test most commonly used is based on the textbook definition of a linear function, namely $f(u+v) = f(u) + f(v)$ (often called the BLR test from [3]): what it means is that we pick $u, v \in \mathbb{F}_2^n$ uniformly at random, compute $u+v$, query the black box to extract $f(u)$, $f(v)$, $f(u+v)$, and check if the aforementioned condition holds. If f passes this test for many pairs (u, v), then f is probably linear. If f fails the test for at least one pair, then f is certainly not linear. We denote by $P_2(f)$ the probability of f failing the test (with probability taken over all pairs $(u, v) \in \mathbb{F}_2^n \times \mathbb{F}_2^n$) and by $\mathrm{d}_{\mathcal{L}}(f)$ the normalized Hamming distance of f to the closest linear function. Several authors have determined upper and lower bounds for $P_2(f)$ as a function of $\mathrm{d}_{\mathcal{L}}(f)$ (see [1, 9] and the references therein).

For cryptographic applications we are not so much interested in whether the function is linear, but rather whether it is affine. For example, such tests play a crucial role in the cube and AIDA attacks (see [6, 12]), which are refined high-order differential attacks, targeted at primitives in stream and block ciphers based on low-degree components. The probabilistic test used in [6] for deciding whether a function f is affine is to check whether $f(u+w) + f(u) + f(w) + f(0) = 0$ holds (for u, w chosen uniformly at random), which can

be viewed as using the BLR test to check whether $f(u) - f(0)$ is linear. The functions of interest f are functions in many variables (typically at least 80 variables), obtained as higher-order derivatives of a function g which describes, for example, the first output bit of the stream cipher as a function of the key and initialisation vector. Although explicit algorithms are available for computing g and f (in the case of the Trivium cipher, the algorithm for computing g starts with some relatively simple functions of algebraic degree two, which are iteratively composed 1152 times for the full cipher, or about 700 times for reduced versions of the cipher), it is not feasible in practice to compute their algebraic normal form, or truth table, or nonlinearity, or Walsh transform. Instead, g is treated as a "black box" function, and f can be evaluated at any given input using several calls to g.

Another test used in the literature for deciding whether a Boolean function is affine is to check whether the equation $f(u + v + w) + f(u) + f(v) + f(w) = 0$ holds, for some $u, v, w \in \mathbb{F}_2^n$ chosen uniformly at random. Like in the case of the linearity test, if f passes the test for many triples u, v, w, then f is probably affine. We denote by $P_3(f)$ the probability of f failing this affinity test (with the probability taken over all triples $(u, v, w) \in \mathbb{F}_2^{3n}$). As in the case of the linearity tests, a natural question is whether $P_3(f)$ is related to $\mathrm{d}_{\mathcal{A}}(f)$, the distance to the closest affine function (note that this is the nonlinearity of f). A lower bound for $P_3(f)$ in terms of $\mathrm{d}_{\mathcal{A}}(f)$ was given in Bellare et al. [1].

A generalization of the tests above was proposed by Zhang and Zheng in [14], where the authors defined the notion of $(k + 1)$-st order nonhomomorphicity of a function f as the probability $P_k(f)$ of failing the test

$$f(u_1 + \cdots + u_k) + f(u_1) + \cdots + f(u_k) = 0,$$

with the probability taken over all tuples $(u_1, \ldots, u_k) \in (\mathbb{F}_2^n)^k$ (see Definition 1). It was shown that for k odd, f is affine if and only if $P_k(f) = 0$; for k even, f is linear if and only if $P_k(f) = 0$; also, still for k even, f is affine if and only if $P_k(f) \in \{0, 1\}$. Furthermore, some bounds on $P_k(f)$ with respect to $\mathrm{d}_{\mathcal{A}}(f)$, for k odd, were given in [14].

In this paper, we firstly improve both the upper and lower bounds presented by Zhang and Zheng in [14] for $P_k(f)$ with k odd (see Sections 3 and 4). Our lower bound holds for arbitrary k and generalizes the lower bound proven in [1] for $k = 2, 3$. The proofs use the techniques employed in [1] as well as additional combinatorial manipulation. We also prove the lower bound we conjectured in [11].

Secondly, we consider the following probabilistic test for affine functions. We can use any probabilistic linearity test, and test whether f is linear or $f + 1$ is linear. If either of these holds, then f is affine. For the nonhomomorphicity test with k even, this is equivalent to testing whether $P_k(f) \in \{0, 1\}$. The fact that f is affine if and only if $P_k(f) \in \{0, 1\}$ was proven in [14]. However, when f is not affine no results were given regarding how the probability of failing this test depends on the nonlinearity of f. In Section 5 we show that upper and lower bounds can be obtained for the value of $\min(P_k(f), P_k(f + 1))$ in the case $k = 2$ (i.e. the BLR test). Namely, using the bounds on failing the BLR linearity test from [1], which depend on the distance to the closest linear function, we show that similar bounds hold for $\min(P_2(f), P_2(f + 1))$, but this time the bounds depend on the distance to the closest affine function. We also show that the refinements of the bounds from [1] given in [9] can be applied to our bounds too.

The nonlinearity of a function f can be estimated by first using any of the above tests and a practical statistical method to estimate the probability of failing that test (as demonstrated in [14]). Then, using (1) or (2), we obtain an estimate for the nonlinearity of f. In Section 6 we analyze the accuracy of the estimation. There are functions f, g such that f has higher probability of failing the test than g, even though f has lower nonlinearity than g. This

was shown in [2] for the BLR test and in [14] for the tests based on the $(k+1)$-st order nonhomomorphicity with k odd. However, the estimates get more accurate as k increases. For example, for $k = 7$, for any given value of $P_7(f)$ we can estimate the nonlinearity as being within an interval of length 0.011 or less if $P_7(f) \leq 0.49$ and length 0.053 or less if $P_7(f) \in [0.49, 0.5]$.

Other nonlinearity tests were proposed for reducing the number of evaluations needed for the black box function, such as [7, 13] (the latter being also useful to estimate the algebraic degree of f). The $(k+1)$-st order nonhomomorphicity for $k = 3$ was used for attacks on actual ciphers in [10]. We intend to push further the connection between the probability of failing these tests and the nonlinearity, as well as look at estimating the nonlinearity of functions of cryptographic interest.

2 Preliminaries

We recall definitions and known results needed for the rest of the paper.

Throughout, n will denote a positive integer. Boolean functions in n variables are functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$, where $\mathbb{F}_2$ is the binary field, and $\mathbb{F}_2^n$ is the n dimensional vector space over $\mathbb{F}_2$. It is well known that any such function can be uniquely represented in its ANF (Algebraic Normal Form), i.e. as a polynomial in $\mathbb{F}_2[x_1, \ldots, x_n]$ of degree at most 1 in each variable. The total degree of the ANF representation is called the *algebraic degree* of f. Functions of algebraic degree at most one are called affine; affine functions with zero constant term are called linear. We will denote by $\mathcal{A}$ the set of affine functions and by $\mathcal{L}$ the set of linear functions in n variables over $\mathbb{F}_2$, if the dimension is understood from the context.

In this paper, like in [1], it will be convenient to use the normalized version of the Hamming distance and weight. More precisely, we define the (normalized) Hamming distance and Hamming weight for vectors $a = (a_1, \ldots, a_t)$ and $b = (b_1, \ldots, b_t)$ in $\mathbb{F}_2^t$, as well as the distance of a vector $a \in \mathbb{F}_2^t$ to a set of vectors $S \subseteq \mathbb{F}_2^t$ as:

$$\begin{aligned} \mathrm{d}(a,b) &= \frac{1}{t}\left|\{i : 1 \leq i \leq t, a_i \neq b_i\}\right|, \\ \mathrm{w}(a) &= \frac{1}{t}\left|\{i : 1 \leq i \leq t, a_i \neq 0\}\right|, \\ \mathrm{d}_S(a) &= \min_{s \in S} \mathrm{d}(a,s). \end{aligned} \tag{3}$$

In the literature, the Hamming weight and distance are more often used without normalization (i.e. in the definitions above, one does not divide by the length of the vector) but we will explain shortly why normalization is useful for our purpose. The truth table of a function f is the vector $TT(f) = (f(v_0), \ldots, f(v_{2^n-1}))$, where v_i are all the elements of $\mathbb{F}_2^n$ in some fixed order, e.g., lexicographical order. The (normalized) Hamming weight, denoted by $\mathrm{w}(f)$, of a Boolean function f is $\mathrm{w}(TT(f))$ and the distance, denoted by $d(f,g)$, between two Boolean functions f, g is $\mathrm{d}(TT(f), TT(g))$.

Of particular importance will be the distance of a function f to the set of affine or of linear functions. The minimum distance to any affine function, $\mathrm{d}_{\mathcal{A}}(f)$, is called the (normalized) *nonlinearity* of f and is a very important cryptographic indicator. It is easy to see that $\mathrm{d}_{\mathcal{A}}(f) = \min(\mathrm{d}_{\mathcal{L}}(f), \mathrm{d}_{\mathcal{L}}(f+1))$. Our motivation for using the normalized version of nonlinearity (based on the normalized version of Hamming distance) is that it allows a meaningful comparison of the nonlinearity of two functions which might not have the same number of variables.

The Fourier-Hadamard transform of a function $f : \mathbb{F}_2^n \to \mathbb{R}$ (the 0/1 values of a Boolean functions are viewed as real numbers for this purpose) is the function $W(f) : \mathbb{F}_2^n \to \mathbb{R}$ defined as

$$W(f)(v) = \frac{1}{2^n} \sum_{u \in \mathbb{F}_2^n} f(u)(-1)^{v \cdot u},$$

where the dot product can be defined as $u \cdot v = \sum_{i=1}^{n} u_i v_i$. Note that we use a normalized version of the transform here. If f is replaced by its sign function, $\hat{f}$, defined by $\hat{f}(u) = (-1)^{f(u)}$, then $W(\hat{f})$ is customarily referred to as the Walsh (or Walsh-Hadamard) transform of f, and the values $W(\hat{f})(v)$ for $v \in \mathbb{F}_2^n$ are called the Walsh coefficients. We will refer to the sequence of output values of the Walsh transform (when the input is ordered lexicographically) as the Walsh spectrum.

We will be using later Parseval's identity (see [5] for example):

$$\sum_{v \in \mathbb{F}_2^n} (W(\hat{f})(v))^2 = 1, \tag{4}$$

which holds for any Boolean function f.

It is well-known [5] and easy to see that the Walsh transform of a Boolean function f expresses its distance to the set of linear functions, and consequently the distance of f to the set of affine functions. Denoting $\ell_a(u) = a \cdot u$, the nonlinearity of f is related to the Walsh transform as follows:

$$\begin{aligned} \mathrm{d}(f, \ell_a) &= \frac{1}{2}\left(1 - W(\hat{f})(a)\right), \\ \mathrm{d}(f, \ell_a + 1) &= \frac{1}{2}\left(1 + W(\hat{f})(a)\right), \\ \mathrm{d}_{\mathcal{L}}(f) &= \frac{1}{2}\left(1 - \max_{v \in \mathbb{F}_2^n} W(\hat{f})(v)\right), \\ \mathrm{d}_{\mathcal{A}}(f) &= \frac{1}{2}\left(1 - \max_{v \in \mathbb{F}_2^n} |W(\hat{f})(v)|\right). \end{aligned} \tag{5}$$

Note that $0 \le \mathrm{d}_{\mathcal{L}}(f) \le \frac{1}{2}$. It is known [5] that $0 \le \mathrm{d}_{\mathcal{A}}(f) \le \frac{1}{2}\left(1 - \frac{1}{2^{\frac{n}{2}}}\right)$. We call a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ ($n \ge 2$) bent if its nonlinearity is exactly $\frac{1}{2}\left(1 - \frac{1}{2^{\frac{n}{2}}}\right)$ (they exist only for even integers n). It is known [5] that f is bent if and only if the absolute values of all of its Walsh coefficients satisfy $|W(\hat{f})(v)| = 2^{-\frac{n}{2}}$.

Definition 1 ([14]) Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function in n variables and let $k \le 2$ be an integer. The $(k+1)$-st order nonhomomorphicity of f, denoted $P_k(f)$, is defined as the probability that the equation $f(u_1 + \cdots + u_k) + f(u_1) + \cdots + f(u_k) = 0$ does not hold, with the probability taken over all tuples $(u_1, \ldots, u_k) \in \mathbb{F}_2^{kn}$ i.e.

$$P_k(f) = \frac{|\{(u_1, \ldots, u_k) \in \mathbb{F}_2^{kn} : f(u_1 + \cdots + u_k) + f(u_1) + \cdots + f(u_k) \neq 0\}|}{2^{kn}}.$$

In other words, $P_k(f)$ is the normalised Hamming weight of the function $F : \mathbb{F}_2^{kn} \to \mathbb{F}_2$, $F(u_1, \ldots, u_k) = f(u_1 + \cdots + u_k) + f(u_1) + \cdots + f(u_k)$. Equivalently, considering $U_1, \ldots, U_k$ independent uniformly distributed random variables in $\mathbb{F}_2^n$, we can define $P_k(f) = P[f(U_1 + \cdots + U_k) + f(U_1) + \cdots + f(U_k) \neq 0]$.

Note that the BLR test corresponds to the particular case of $k = 2$.

3 Improved bounds on the probability of failure of existing affinity tests

We consider the test of whether a function is affine by checking whether $f(u_1+\cdots+u_k)+f(u_1)+\cdots+f(u_k)=0$, for some fixed odd integer k. We examine the relationship between the probability $P_k(f)$ of failing this test and $\mathrm{d}_{\mathcal{A}}(f)$, the nonlinearity of f. It is well known, and easy to prove, that f is affine if and only if $P_k(f)=0$.

A lower bound for $P_3(f)$ was proven in [1, Lemma 5.1] (with $x=\mathrm{d}_{\mathcal{A}}(f)$):

$$\begin{aligned} P_3(f) &\geq \max\left(8x(1-x)\left(\frac{1}{2}-x\right), 2x(1-x)\right) \\ &= \begin{cases} 8x(1-x)\left(\frac{1}{2}-x\right) & \text{if } x \leq \frac{1}{4} \\ 2x(1-x) & \text{if } x > \frac{1}{4}. \end{cases} \end{aligned} \tag{6}$$

The following lower and upper bounds were given in [14] for k odd (we reformulated them to use the normalized version):

$$\frac{1}{2}\left(1-2^n\left(1-2\mathrm{d}_{\mathcal{A}}(f)\right)^{k+1}\right) \leq P_k(f) \leq \frac{1}{2}\left(1-\frac{1}{2^{(k-1)n/2}}\right). \tag{7}$$

We improve on the bounds (7) as follows:

Theorem 2 *Let $f:\mathbb{F}_2^n \to \mathbb{F}_2$ and let $k \geq 2$ be an integer. Then:*

$$\frac{1}{2}\left(1-(1-2x)^{k-1}\right) \leq P_k(f), \tag{8}$$

where $x=\mathrm{d}_{\mathcal{A}}(f)$ if k is odd, and $x=\mathrm{d}_{\mathcal{L}}(f)$ if k is even.

For k odd we have the upper bound $P_k(f) \leq \mathrm{Upper}_k(\mathrm{d}_{\mathcal{A}}(f))$, where

$$Upper_k(x) = \frac{1}{2}\left(1-(1-2x)^{k+1}\right). \tag{9}$$

If we allow the bound to also depend on n, we have the improved bound $P_k(f) \leq \mathrm{Upper}_{n,k}(\mathrm{d}_{\mathcal{A}}(f))$, where

$$Upper_{n,k}(x) = \frac{1}{2}\left(1-(1-2x)^{k+1}-\frac{1}{(2^n-1)^{\frac{k-1}{2}}}(4x(1-x))^{\frac{k+1}{2}}\right). \tag{10}$$

Proof In [8, Theorem 3.1], [14, Theorem 2] (and, for $k \leq 3$, in [1]) the following expression for P_k is obtained (we reformulate it for the normalized versions of the Walsh transform and nonhomomorphicity):

$$P_k(f) = \frac{1}{2}\left(1-\sum_{u\in\mathbb{F}_2^n}(W(\hat{f})(u))^{k+1}\right), \tag{11}$$

where $\hat{f}(x)=(-1)^{f(x)}$ and $W(\hat{f})$ is the Walsh transform of f.

In order to obtain the lower bound in the statement we need an upper bound on the sum $\sum_{u\in\mathbb{F}_2^n}(W(\hat{f})(u))^{k+1}$, which we obtain by a technique similar to the one of [1]:

$$\sum_{u\in\mathbb{F}_2^n}\left(W(\hat{f})(u)\right)^{k+1} \le \max_{u\in\mathbb{F}_2^n}\left(W(\hat{f})(u)\right)^{k-1}\sum_{u\in\mathbb{F}_2^n}(W(\hat{f})(u))^2 = \max_{u\in\mathbb{F}_2^n}\left(W(\hat{f})(u)\right)^{k-1},$$

where the last equality uses Parseval's identity (4). Using (5), we obtain (8) as follows:

$$\begin{aligned}\max_{u\in\mathbb{F}_2^n}\left(W(\hat{f})(u)\right)^{k-1} &= \begin{cases}\left(\max_{u\in\mathbb{F}_2^n} W(\hat{f})(u)\right)^{k-1} & \text{if } k \text{ is even}\\ \left(\max_{u\in\mathbb{F}_2^n} |W(\hat{f})(u)|\right)^{k-1} & \text{if } k \text{ is odd}\end{cases}\\ &= \begin{cases}(1-2\mathrm{d}_{\mathcal{L}}(f))^{k-1} & \text{if } k \text{ is even}\\ (1-2\mathrm{d}_{\mathcal{A}}(f))^{k-1} & \text{if } k \text{ is odd.}\end{cases}\end{aligned}$$

When k is odd (so the exponent $k+1$ is even), all the terms in the sum in (11) are non-negative, so a simple lower bound for this sum is

$$\sum_{u\in\mathbb{F}_2^n}\left(W(\hat{f})(u)\right)^{k+1} \ge \max_{u\in\mathbb{F}_2^n}\left(W(\hat{f})(u)\right)^{k+1} = \left(\max_{u\in\mathbb{F}_2^n}|W(\hat{f})(u)|\right)^{k+1} = (1-2\mathrm{d}_{\mathcal{A}}(f))^{k+1},$$

which gives the upper bound (9). For the bound (10), let $u_0\in\mathbb{F}_2^n$ be such that $|W(\hat{f})(u_0)| = \max_{u\in\mathbb{F}_2^n}|W(\hat{f})(u)|$. We have

$$\sum_{u\in\mathbb{F}_2^n}\left(W(\hat{f})(u)\right)^{k+1} = |W(\hat{f})(u_0)|^{k+1} + \sum_{u\in\mathbb{F}_2^n\setminus\{u_0\}}\left(\left(W(\hat{f})(u)\right)^2\right)^{\frac{k+1}{2}}.$$

Recall that the weighted power means inequality states that for any integers $m\ge 1$, $j\ge 2$ and any positive real numbers $a_1,\ldots,a_m$ we have $\sum_{i=1}^m \frac{1}{m}a_i^j \ge \left(\frac{1}{m}\sum_{i=1}^m a_i\right)^j$ (see for example [4, Chapter III]). Using this inequality and Parseval's identity, we obtain

$$\begin{aligned}\sum_{u\in\mathbb{F}_2^n\setminus\{u_0\}}\left(\left(W(\hat{f})(u)\right)^2\right)^{\frac{k+1}{2}} &\ge \frac{2^n-1}{(2^n-1)^{\frac{k+1}{2}}}\left(\sum_{u\in\mathbb{F}_2^n\setminus\{u_0\}}(W(\hat{f})(u))^2\right)^{\frac{k+1}{2}}\\ &= \frac{1}{(2^n-1)^{\frac{k-1}{2}}}\left(1-|W(\hat{f})(u_0)|^2\right)^{\frac{k+1}{2}}.\end{aligned}$$

Substituting $|W(\hat{f})(u_0)| = 1-2\mathrm{d}_{\mathcal{A}}(f)$, we obtain (10). □

Note that for $k\ge 3$ odd, the bounds in the theorem above are better than the bounds (7). The lower bound in (7) is negative when $\mathrm{d}_{\mathcal{A}}(f) < \frac{1}{2}\left(1-\frac{1}{2^{\frac{n}{k+1}}}\right)$, so it does not provide any useful information in that range. When it is positive, it is still always smaller than the lower bound in (8), only reaching equality when $\mathrm{d}_{\mathcal{A}}(f)$ attains its maximum value, namely $\frac{1}{2}\left(1-\frac{1}{2^{\frac{n}{2}}}\right)$. The upper bound in (7) does not depend on $\mathrm{d}_{\mathcal{A}}(f)$, whereas the one in (10) increases continuously from 0 to $\frac{1}{2}\left(1-\frac{1}{2^{(k-1)n/2}}\right)$ as $\mathrm{d}_{\mathcal{A}}(f)$ increases from 0 to $\frac{1}{2}\left(1-\frac{1}{2^{\frac{n}{2}}}\right)$.

We examine the tightness of the bounds in Theorem 2. The upper bound (9) cannot be reached (except for the trivial case $\mathrm{d}_{\mathcal{A}}(f) = 0$ and $P_k(f) = 0$) because $\mathrm{Upper}_{n,k}(x) < \mathrm{Upper}_k(x)$ for $0 < x < 0.5$. Note however that $\mathrm{Upper}_k(x)$ is the limit of $\mathrm{Upper}_{n,k}(x)$, as n approaches infinity. We found experimentally functions f for which $P_k(f)$ is very close to the upper bound $\mathrm{Upper}_k(\mathrm{d}_{\mathcal{A}}(f))$, while $\mathrm{d}_{\mathcal{A}}(f)$ covers many values throughout the interval $(0, 0.5)$, see the last graph in the Appendix. We suspect therefore that this upper bound cannot be improved much (as a bound which is independent of n).

The examples below present functions for which the upper bound (10) as well as the lower bound in Theorem 2 are attained.

Example 3 For n even and k odd, consider a bent function in n variables, for example $f(x_1, \ldots, x_n) = x_1x_2 + x_3x_4 + \cdots + x_{n-1}x_n$. The nonlinearity achieves the maximum possible value for a function in n variables, namely $\mathrm{d}_{\mathcal{A}}(f) = \frac{1}{2}\left(1 - \frac{1}{2^{\frac{n}{2}}}\right)$. All the Walsh coefficients of a bent function are equal to $\pm\frac{1}{2^{\frac{n}{2}}}$, with $2^{n-1} + 2^{\frac{n}{2}-1}$ having one sign and $2^{n-1} - 2^{\frac{n}{2}-1}$ the opposite sign. Using (11) we can compute

$$\begin{aligned} P_k(f) &= \frac{1}{2}\left(1 - 2^n\left(\frac{1}{2^{\frac{n}{2}}}\right)^{k+1}\right) \\ &= \frac{1}{2}\left(1 - \left(\frac{1}{2^{\frac{n}{2}}}\right)^{k-1}\right) \\ &= \frac{1}{2}\left(1 - (1 - 2\mathrm{d}_{\mathcal{A}}(f))^{k-1}\right). \end{aligned}$$

Note that in this case

$$\frac{1}{2}\left(1 - (1 - 2\mathrm{d}_{\mathcal{A}}(f))^{k-1}\right) = P_k(f) = \mathrm{Upper}_{n,k}(\mathrm{d}_{\mathcal{A}}(f))$$

so both the lower bound (8) and the upper bound (10) are attained.

Example 4 Consider the function $f(x_1, \ldots, x_n) = x_1x_2\cdots x_n$ with $n \geq 2$. The Walsh coefficients can be easily computed directly from the definition:

$$W(\hat{f})(u_1, \ldots, u_n) = \begin{cases} 1 - \frac{1}{2^{n-1}} & \text{if } (u_1, \ldots, u_n) = (0, \ldots, 0) \\ \frac{1}{2^{n-1}}(-1)^{1+\sum_{i=1}^n u_i} & \text{otherwise.} \end{cases}$$

The nonlinearity is $\mathrm{d}_{\mathcal{A}}(f) = \frac{1}{2^n}$. Using (11) we have for k odd

$$\begin{aligned} P_k(f) &= \frac{1}{2}\left(1 - \left(1 - \frac{1}{2^{n-1}}\right)^{k+1} - (2^n - 1)\left(\frac{1}{2^{n-1}}\right)^{k+1}\right) \\ &= \frac{1}{2}\left(1 - \frac{(2^{n-1} - 1)^{k+1} + (2^n - 1)}{2^{(n-1)(k+1)}}\right). \end{aligned}$$

One can verify that in this case $P_k(f) = \mathrm{Upper}_{k,n}(\mathrm{d}_{\mathcal{A}}(f))$ so the upper bound (10) is attained.

Example 5 Consider an arbitrary function in m variables, $f'(x_1, \ldots, x_m)$. We can view it as a function in a larger number n of variables for any $n \geq m$ by defining $f(x_1, \ldots, x_n) =$

$f'(x_1, \ldots, x_m)$. We show that f and f' have the same nonlinearity and $P_k(f) = P_k(f')$. To this end, we examine the Walsh transform. Denoting $x' = (x_1, \ldots, x_m)$ and $x'' = (x_{m+1}, \ldots, x_n)$, as well as, $y' = (y_1, \ldots, y_m)$ and $y'' = (y_{m+1}, \ldots, y_n)$, we have

$$\begin{aligned} W(\hat{f})(x_1, \ldots, x_n) &= \frac{1}{2^n} \sum_{(y', y'') \in \mathbb{F}_2^n} (-1)^{f(y', y'') + x' \cdot y' + x'' \cdot y''} \\ &= \left(\frac{1}{2^{n-m}} \sum_{y'' \in \mathbb{F}_2^{n-m}} (-1)^{x'' \cdot y''} \right) \left(\frac{1}{2^m} \sum_{y' \in \mathbb{F}_2^m} (-1)^{f'(y') + x' \cdot y'} \right) \\ &= W(\hat{f'})(x') \frac{1}{2^{n-m}} \sum_{y'' \in \mathbb{F}_2^{n-m}} (-1)^{x'' \cdot y''} \\ &= \begin{cases} W(\hat{f'})(x') & \text{if } x'' = 0 \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{12}$$

by using [5, Lemma 2.9]. Therefore we conclude that $\mathrm{d}_{\mathcal{A}}(f) = \mathrm{d}_{\mathcal{A}}(f')$ using (5); also $P_k(f) = P_k(f')$ using (11).

We consider now the function $f(x_1, \ldots, x_n) = x_1x_2 + x_3x_4 + \cdots + x_{m-1}x_m$ with m even and $m \leq n$ and let k be odd. Using the argument above and the computation in Example 3 we know that $\mathrm{d}_{\mathcal{A}}(f) = \frac{1}{2}\left(1 - \frac{1}{2^{\frac{m}{2}}}\right)$ and $P_k(f) = \frac{1}{2}\left(1 - (1 - 2\mathrm{d}_{\mathcal{A}}(f))^{k-1}\right)$, so this function reaches the lower bound in Theorem 2 as well.

Summarising the examples above, for each fixed number of variables n and each odd k, the upper bound $\mathrm{Upper}_{n,k}$ in Theorem 2 is reached at nonlinearity $\frac{1}{2^n}$ (which is the lowest possible non-zero nonlinearity) and if n is even, also at $\frac{1}{2}\left(1 - \frac{1}{2^{\frac{n}{2}}}\right)$ (which is the highest possible nonlinearity). The lower bound in Theorem 2 is reached for nonlinearities $\frac{1}{2^n}$ as well as all nonlinearities of the form $\frac{1}{2}\left(1 - \frac{1}{2^{\frac{m}{2}}}\right)$ for m even, $m \leq n$, i.e. $\frac{1}{4}, \frac{3}{8}, \frac{7}{16}, \ldots$, for $m = 2, 4, 6, \ldots$, respectively. In between these values, the lower bound might not be tight. Indeed for $\mathrm{d}_{\mathcal{A}}(f) < \frac{1}{4}$, (6) provides a better lower bound for $k = 3$.

We conjectured in [11] that the lower bound (6) can be generalized to arbitrary odd $k \geq 3$, as follows:

Conjecture 6 ([11]) *For any odd $k \geq 3$, putting $x = \mathrm{d}_{\mathcal{A}}(f)$, we have*

$$\begin{aligned} P_k(f) &\geq \frac{1}{2} \max\left(1 - (1-2x)^{k+1} - 2^{k+1}x^k(1-x),\ 1 - (1-2x)^{k-1}\right) \\ &= \begin{cases} \frac{1}{2}\left(1 - (1-2x)^{k+1} - 2^{k+1}x^k(1-x)\right) & \text{if } x \leq \frac{1}{4} \\ \frac{1}{2}\left(1 - (1-2x)^{k-1}\right) & \text{if } x > \frac{1}{4}. \end{cases} \end{aligned} \tag{13}$$

In the next section we will prove this conjecture.

4 Reformulated conjecture and its proof

Theorem 7 *Let f be a Boolean function in n variables, $k \geq 2$ an integer and ℓ a linear function in n variables if k is even or an affine function if k is odd. Then*

$$P_k(f) = \frac{1}{2}\left(1 - (1 - 2\mathrm{d}(f,\ell))^{k+1} + (-1)^{k+1}2^{k+1}\left(\mathrm{d}(f,\ell)^{k+1} - sl(f,\ell)\right)\right), \qquad (14)$$

where for any Boolean function h in n variables $sl(f,h)$ (called the "slack" in [1]) is defined as

$$sl(f,h) = P\left(f(u_1) \neq h(u_1), \ldots, f(u_k) \neq h(u_k), f\left(\sum_{i=1}^{k} u_i\right) \neq h\left(\sum_{i=1}^{k} u_i\right)\right)$$

with the probability taken over all $u_1, \ldots, u_k \in \mathbb{F}_2^n$.

Proof The first part of the proof follows the lines of [1, Lemma 2.3]. Denote $x = \mathrm{d}(f,\ell)$. Let $u_1, \ldots, u_k \in \mathbb{F}_2^n$ and denote $u_{k+1} = \sum_{i=1}^{k} u_i$. The function f fails the test $f(u_1 + \cdots + u_k) + f(u_1) + \cdots + f(u_k) = 0$ exactly for those values $u_1, \ldots, u_k$ for which an odd number of the values $f(u_1) - \ell(u_1), \ldots, f(u_{k+1}) - \ell(u_{k+1})$ are equal to 1 (note that ℓ always passes the test). Denote by A_j the probability that the first j of these $k+1$ values are equal to 1 and the rest are equal to 0, i.e.

$$A_j := P\left(g(u_1) = 1, \ldots, g(u_j) = 1, g(u_{j+1}) = 0, \ldots, g(u_{k+1}) = 0\right),$$

where, for ease of notation, we denoted $g = f - \ell$, and the probability is taken over all the 2^{kn} elements of the set $V = \{(u_1, \ldots, u_{k+1}) \in (\mathbb{F}_2^n)^{k+1} : \sum_{i=1}^{k+1} u_i = 0\}$. For any subset $I \subseteq \{1, \ldots, k+1\}$ of cardinality j, A_j also equals the probability (again over all $(u_1, \ldots, u_{k+1}) \in V$) that $g(u_i) = 1$ for all $i \in I$ and $g(u_i) = 0$ for all $i \in \{1, \ldots, k+1\} \setminus I$. Therefore, taking into account that there are $\binom{k+1}{j}$ subsets of cardinality j, we obtain

$$P_k(f) = \sum_{\substack{1 \leq j \leq k+1 \\ j \ odd}} \binom{k+1}{j} A_j.$$

For each fixed i, the probability $P(f(u_i) - \ell(u_i) = 1)$ over all $u_i \in \mathbb{F}_2^n$ equals $x = \mathrm{d}(f,\ell)$. For any j with $0 \leq j \leq k$ we have

$$\begin{aligned} A_j &= P\left(g(u_1) = 1, \ldots, g(u_j) = 1, g(u_{j+1}) = 0, \ldots, g(u_k) = 0\right) \\ &\quad - P\left(g(u_1) = 1, \ldots, g(u_j) = 1, g(u_{j+1}) = 0, \ldots, g(u_k) = 0, g(u_{k+1}) = 1\right) \\ &= x^j(1-x)^{k-j} - A_{j+1} = \cdots \\ &= \sum_{i=j}^{k} (-1)^{i-j} x^i (1-x)^{k-i} + (-1)^{k-j+1} A_{k+1}. \end{aligned}$$

We obtain

$$P_k(f) = \sum_{\substack{1\le j\le k+1\\ j\ odd}} \binom{k+1}{j} \sum_{i=j}^{k} (-1)^{i-j} x^i (1-x)^{k-i} + A_{k+1} \sum_{\substack{1\le j\le k+1\\ j\ odd}} (-1)^{k-j+1} \binom{k+1}{j}. \tag{15}$$

In the second part of the proof we will obtain a closed form for the formula above. Firstly, let us process the inner sum in (15); we replace the index of summation by $u = i - j$ and then use the well-known identity $a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + \cdots + b^{n-1})$:

$$\begin{aligned}
\sum_{i=j}^{k} (-1)^{i-j} x^i (1-x)^{k-i} &= \sum_{u=0}^{k-j} (-1)^u x^{u+j} (1-x)^{k-j-u} \\
&= x^j \sum_{u=0}^{k-j} (-x)^u (1-x)^{k-j-u} \\
&= x^j \frac{(1-x)^{k+1-j} - (-x)^{k+1-j}}{(1-x)-(-x)} \\
&= x^j (1-x)^{k+1-j} - (-1)^{k+1-j} x^{k+1}.
\end{aligned}$$

Substituting this in (15) and since $(-1)^{k+1-j} = -(-1)^{k+1}$, when j is odd, we obtain

$$P_k(f) = \sum_{\substack{1\le j\le k+1\\ j\ odd}} \binom{k+1}{j} x^j (1-x)^{k+1-j} + (-1)^{k+1}(x^{k+1} - A_{k+1}) \sum_{\substack{1\le j\le k+1\\ j\ odd}} \binom{k+1}{j}. \tag{16}$$

We note that the first sum consists of alternating terms of a binomial expansion. The following result is therefore useful:

$$\sum_{\substack{0\le j\le m\\ j\ odd}} \binom{m}{j} a^j b^{m-j} = \frac{1}{2}\left((a+b)^m - (-a+b)^m\right), \tag{17}$$

where $m \ge 1$ is an integer and a, b indeterminates. This is a known result, but for a quick proof, we denote by A and B the following quantities

$$\begin{aligned}
A &= \sum_{\substack{0\le j\le m\\ j\ odd}} \binom{m}{j} a^j b^{m-j} = -\sum_{\substack{0\le j\le m\\ j\ odd}} \binom{m}{j} (-a)^j b^{m-j}, \\
B &= \sum_{\substack{0\le j\le m\\ j\ even}} \binom{m}{j} a^j b^{m-j} = \sum_{\substack{0\le j\le m\\ j\ even}} \binom{m}{j} (-a)^j b^{m-j},
\end{aligned}$$

and use the fact that $A + B = (a+b)^m$ and $-A + B = (-a+b)^m$. For $a = b = 1$, (17) becomes

$$\sum_{\substack{0\le j\le m\\ j\ odd}} \binom{m}{j} = 2^{m-1}. \tag{18}$$

Using Equations (17), (18) and $A_{k+1} = sl(f, \ell)$ in (16), we obtain (14):

$$\begin{aligned} P_k(f) &= \frac{1}{2}\left((x+1-x)^{k+1} - (-x+1-x)^{k+1}\right) + (-1)^{k+1}2^k(x^{k+1} - A_{k+1}) \\ &= \frac{1}{2}\left(1-(1-2x)^{k+1} + (-1)^{k+1}2^{k+1}(x^{k+1} - sl(f,\ell))\right). \end{aligned}$$

This concludes the proof. □

We are now ready to prove Conjecture 6; we will, in fact, prove a more general result that also includes the case of k even.

Corollary 8 *For odd k we have $P_k(f) \geq \mathrm{Lower}_k(\mathrm{d}_{\mathcal{A}}(f))$ and for even k we have $P_k(f) \geq \mathrm{Lower}_k(\mathrm{d}_{\mathcal{L}}(f))$ where $\mathrm{Lower}_k(x)$ equals*

$$\begin{cases} \frac{1}{2}\max\left(1-(1-2x)^{k+1} - 2^{k+1}x^k(1-x),\, 1-(1-2x)^{k-1}\right) & \text{if } k \text{ is odd} \\ \frac{1}{2}\max\left(\left(1-(1-2x)^{k+1} - 2^{k+1}x^{k+1}\right),\, 1-(1-2x)^{k-1}\right) & \text{if } k \text{ is even.} \end{cases}$$

In more detail, for k odd we have

$$\mathrm{Lower}_k(x) = \begin{cases} \frac{1}{2}\left(1-(1-2x)^{k+1} - 2^{k+1}x^k(1-x)\right) & \text{if } x < \frac{1}{4} \\ \frac{1}{2}\left(1-(1-2x)^{k-1}\right) & \text{if } x \geq \frac{1}{4}. \end{cases}$$

Proof Put $H_k(x) = \frac{1}{2}\left(1-(1-2x)^{k-1}\right)$ and

$$G_k(x) = \begin{cases} \frac{1}{2}\left(1-(1-2x)^{k+1} - 2^{k+1}x^k(1-x)\right) & \text{if } k \text{ is odd} \\ \frac{1}{2}\left(\left(1-(1-2x)^{k+1} - 2^{k+1}x^{k+1}\right)\right) & \text{if } k \text{ is even.} \end{cases}$$

The $H_k(x)$ component of the lower bound was proven in Theorem 2, so we concentrate on the $G_k(x)$ component.

For k even, Theorem 7 gives

$$P_k(f) = \frac{1}{2}\left(1-(1-2\mathrm{d}(f,\ell))^{k+1} - 2^{k+1}\mathrm{d}(f,\ell)^{k+1} + 2^{k+1}sl(f,\ell)\right) \tag{19}$$

for any linear function ℓ. Using the fact that $sl(f, \ell) \geq 0$ and choosing ℓ to be a linear function whose distance to f is minimal, we obtain $P_k(f) \geq G_k(\mathrm{d}_{\mathcal{L}}(f))$ as required.

For any function h (affine or not) we have

$$sl(f,h) \leq P\left(f(u_1) \neq h(u_1), \ldots, f(u_k) \neq h(u_k)\right) = \mathrm{d}(f,h)^k$$

with the probability taken over all tuples $(u_1, \ldots, u_k) \in (\mathbb{F}_2^n)^k$. Combining this inequality with Theorem 7 for k odd gives

$$P_k(f) \geq \frac{1}{2}\left(1-(1-2\mathrm{d}(f,\ell))^{k+1} + 2^{k+1}\left(\mathrm{d}(f,\ell)^{k+1} - \mathrm{d}(f,\ell)^k\right)\right) \tag{20}$$

for any affine function ℓ. Choosing ℓ to be an affine function whose distance to f is minimal, we obtain $P_k(f) \geq G_k(\mathrm{d}_{\mathcal{A}}(f))$ as required.

Surely, we can ask ourselves whether it is possible that another affine/linear function (for k odd/even) say ℓ_1, which is further from f, i.e. $\mathrm{d}(f, \ell_1) > \mathrm{d}(f, \ell_0)$, could yield a better lower bound, i.e. $G_k(\mathrm{d}(f, \ell_1)) > G_k(\mathrm{d}(f, \ell_0))$. This is *not* the case, and we give a sketch of the proof for k odd. Firstly, the reader can verify that $G_k(x) \geq 0$ on $[0, 0.5]$, $G_k(x) \leq 0$ on $[0.5, 1]$, and that on the interval $[0.25, 0.5]$, the function G_k is monotonically decreasing. Therefore, when $\mathrm{d}(f, \ell_0) \geq 0.25$, keeping in mind that $0 \leq \mathrm{d}(f, \ell_0) < 0.5$ and

$\mathrm{d}(f,\ell_0) < \mathrm{d}(f,\ell_1) \leq 1$, we have indeed $G_k(\mathrm{d}(f,\ell_1)) \leq G_k(\mathrm{d}(f,\ell_0))$. Secondly, when $\mathrm{d}(f,\ell_0) < 0.25$, the triangle inequality gives

$$\mathrm{d}(f,\ell_1) \geq \mathrm{d}(\ell_1,\ell_0) - \mathrm{d}(f,\ell_0) \geq 0.5 - \mathrm{d}(f,\ell_0) \geq 0.25.$$

Therefore $G_k(\mathrm{d}(f,\ell_1)) \leq G_k(0.5 - \mathrm{d}(f,\ell_0))$ by the same argument as above. To show that $G_k(0.5 - \mathrm{d}(f,\ell_0)) \leq G_k(\mathrm{d}(f,\ell_0))$ we compute

$$G_k(x) - G_k(0.5 - x) = 2x(1-2x)((1-2x)^{k-1} - (2x)^{k-1}),$$

which is greater than or equal to zero on $[0, 0.25]$.

Finally, the more explicit expression (19) for k odd is obtained by verifying that $G_k(x) > H_k(x)$ when $x \in \left[0, \frac{1}{4}\right)$, $G_k(x) < H_k(x)$ when $x \in \left[\frac{1}{4}, \frac{1}{2}\right]$, and $G_k(\frac{1}{4}) = H_k(\frac{1}{4})$. A similar situation happens for k even, but the intersection of the two functions does not occur at $\frac{1}{4}$, but at a point whose value depends on k, and is in the interval $\left[\frac{1}{4}, \frac{1}{3}\right]$. □

The new lower bound in Corollary 8 is attained for k odd by some functions with nonlinearity in the range $0 < \mathrm{d}_{\mathcal{A}}(f) < \frac{1}{4}$ and for k even by some functions with $0 < \mathrm{d}_{\mathcal{L}}(f) < \frac{1}{4}$:

Example 9 Let $f(x_1, x_2, \ldots, x_n) = x_1 x_2 \cdots x_m$ with $3 \leq m \leq n$. Using the computations in Example 4 and the same arguments as in Example 5 we see that $\mathrm{d}_{\mathcal{A}}(f) = \mathrm{d}_{\mathcal{L}}(f) = \frac{1}{2^m}$ and the Walsh spectrum consists of one element equal to $1 - \frac{1}{2^{m-1}}$, 2^{m-1} elements equal to $\frac{1}{2^{m-1}}$, and $2^{m-1} - 1$ elements equal to $-\frac{1}{2^{m-1}}$, the remaining elements being zero.

For k odd, like in Example 4 we compute

$$P_k(f) = \frac{1}{2}\left(1 - \frac{(2^{m-1}-1)^{k+1} + (2^m - 1)}{2^{(m-1)(k+1)}}\right).$$

On the other hand, computing $\mathrm{Lower}_k(x)$ defined in Corollary 8 for $x = \mathrm{d}_{\mathcal{A}}(f) = \frac{1}{2^m}$ we obtain

$$\begin{aligned}\mathrm{Lower}_k(\mathrm{d}_{\mathcal{A}}(f)) &= \frac{1}{2}\left(1 - \left(1 - \frac{1}{2^{m-1}}\right)^{k+1} - 2^{k+1}\frac{1}{2^{mk}}\left(1 - \frac{1}{2^m}\right)\right)\\ &= \frac{1}{2}\left(1 - \frac{(2^{m-1}-1)^{k+1} + (2^m-1)}{2^{(m-1)(k+1)}}\right),\end{aligned}$$

so the lower bound is attained.

For k even we compute using (11)

$$\begin{aligned}P_k(f) &= \frac{1}{2}\left(1 - \left(1 - \frac{1}{2^{m-1}}\right)^{k+1} - 2^{m-1}\left(\frac{1}{2^{m-1}}\right)^{k+1} + (2^{m-1}-1)\left(\frac{1}{2^{m-1}}\right)^{k+1}\right)\\ &= \frac{1}{2}\left(1 - \left(1 - \frac{1}{2^{m-1}}\right)^{k+1} - \left(\frac{1}{2^{m-1}}\right)^{k+1}\right)\\ &= \mathrm{Lower}_k\left(\frac{1}{2^m}\right),\end{aligned}$$

so again the lower bound is attained.

This shows that for each fixed n our lower bound is attained at nonlinearity (for k odd) or $\mathrm{d}_{\mathcal{L}}(f)$ (for k even) equal to $\frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \ldots, \frac{1}{2^n}$, but in between these values, the bound might not be tight.

While not the purpose of this paper, we get an interesting consequence of the previous theorem, namely an upper bound for the moments of the Walsh coefficients.

Corollary 10 *For any integer $k \geq 2$, the $(k+1)$-st moments of the Walsh transform satisfy*

$$\sum_{u \in \mathbb{F}_2^n} (W(\hat{f})(u))^{k+1} \leq \begin{cases} \min(y^{k+1} + (1-y)^k(1+y),\, y^{k-1}) & \text{if } k \text{ is odd} \\ \min(y^{k+1} + (1-y)^{k+1},\, y^{k-1}) & \text{if } k \text{ is even,} \end{cases}$$

where $y = \max_{u \in \mathbb{F}_2^n} |W(\hat{f})(u)|$ for k odd and $y = \max_{u \in \mathbb{F}_2^n} (W(\hat{f})(u))$ for k even.

Proof The claim follows by using (5), (11) and the previous corollary. □

5 Affinity tests using linearity tests and bounds on the probability of failure

In this section we focus on the test $f(u_1 + \cdots + u_k) + f(u_1) + \cdots + f(u_k) = 0$ for k even, and on the probability $P_k(f)$ of failing this test. It is shown in [14] that f is linear if and only if $P_k(f) = 0$; moreover, f is affine if and only if $P_k(f) \in \{0, 1\}$. The test can therefore be used as a probabilistic affinity test as follows: run the test several times on f, and if f always passes the test (suggesting a probability of failure $P_k(f) = 0$), or f always fails the test (suggesting that $P_k(f) = 1$), then declare f to be affine. Note that $f+1$ passes the linearity test above for some given tuples if and only if f fails the test for those same tuples; therefore we have $P_k(f+1) = 1 - P_k(f)$. Another way of looking at this affinity test is that we are testing both f and $f+1$ for linearity, and if one of them passes all the tests and is declared linear then f can be declared affine. Any other linearity test could be used this way as an affinity test.

When f is not affine however (and therefore neither f nor $f+1$ are linear), there are to our knowledge no results regarding the relationship of the probability $P_k(f)$ of failing the test (for k even) and the nonlinearity $\mathrm{d}_{\mathcal{A}}(f)$ of f; the existing lower and upper bounds on $P_k(f)$ depend on $\mathrm{d}_{\mathcal{L}}(f)$, the distance of f to the set of linear functions. Since this affinity test is equivalent to testing both f and $f+1$ for linearity, it seems natural to consider both $P_k(f)$ and $P_k(f+1)$ when examining a connection to $\mathrm{d}_{\mathcal{A}}(f)$. We define

$$\overline{P}_k(f) := \min(P_k(f), P_k(f+1)) = \min(P_k(f), 1 - P_k(f)).$$

and study its relation to $\mathrm{d}_{\mathcal{A}}(f)$. Further motivation for this choice is given in Remark 12. For $k = 2$, we will prove lower and upper bounds for $\overline{P}_2(f)$ in terms of the nonlinearity of f.

In Bellare et al. [1] lower and upper bounds were given for $P_2(f)$ in terms of $\mathrm{d}_{\mathcal{L}}(f)$. Namely, it was proven that

$$\mathrm{Lower}_2(\mathrm{d}_{\mathcal{L}}(f)) \leq P_2(f) \leq \mathrm{Upper}_2(\mathrm{d}_{\mathcal{L}}(f))), \tag{21}$$

where $\mathrm{Lower}_2, \mathrm{Upper}_2 : [0, \frac{1}{2}] \to \mathbb{R}$. The function $\mathrm{Lower}_2(x)$ is defined as

$$\mathrm{Lower}_2(x) = \begin{cases} 3x - 6x^2 & \text{if } 0 \leq x \leq \frac{5}{16} \\ \frac{45}{128} & \text{if } \frac{5}{16} \leq x \leq \frac{45}{128} \\ x & \text{if } \frac{45}{128} \leq x \leq \frac{1}{2} \end{cases} \tag{22}$$

(observe that $\frac{3}{16}$, $\frac{5}{16}$ are the two solutions of the equation $3x - 6x^2 = \frac{45}{128}$, and so, since $\frac{45}{128} > \frac{5}{16}$, then the value $\text{Lower}_2(x) = \frac{45}{128}$ is greater than $3x - 6x^2$ on the interval $\frac{5}{16} \leq x \leq \frac{45}{128}$). The function $\text{Upper}_2(x)$ is defined as $\text{Upper}_2(0) = 0$ and for $x > 0$

$$\text{Upper}_2(x) = 3x - 6x^2 + 2^{2\lfloor \log_2 x \rfloor + 2} + 12\left(x - 2^{\lfloor \log_2 x \rfloor}\right)^2. \tag{23}$$

We now prove bounds for $\overline{P}_2(f)$ in terms of $\text{d}_{\mathcal{A}}(f)$, the distance to the closest *affine* function, which is the natural parameter to consider when testing if a function is affine. Note that in the following theorem, although the bounds look similar to the bounds in (21) above, there is a subtle and important difference: the bounds are now a function of $\text{d}_{\mathcal{A}}(f)$, the distance to the closest *affine* function, whereas in (21) the bounds are expressed in terms of $\text{d}_{\mathcal{L}}(f)$, the distance to the closest *linear* function.

Theorem 11 *Let $\overline{P}_2(f) = \min(P_2(f), 1 - P_2(f))$, where $P_2(f)$ is the probability of failure of the BLR test. We have*

$$\text{Lower}_2\left(\text{d}_{\mathcal{A}}(f)\right) \leq \overline{P}_2(f) \leq \min\left(\frac{1}{2}, \text{Upper}_2(\text{d}_{\mathcal{A}}(f))\right), \tag{24}$$

where $\text{d}_{\mathcal{A}}(f)$ is the nonlinearity of f and $\text{Lower}_2(x)$, $\text{Upper}_2(x)$ are as defined above in (22), respectively, (23).

Proof We know that $\text{d}_{\mathcal{A}}(f) = \min(\text{d}_{\mathcal{L}}(f), \text{d}_{\mathcal{L}}(f+1))$. We can assume, without loss of generality, that $\text{d}_{\mathcal{L}}(f) \leq \text{d}_{\mathcal{L}}(f+1)$ (otherwise, we can just replace f by $f+1$, and $\overline{P}_2(f)$ is unchanged) and therefore $\text{d}_{\mathcal{A}}(f) = \text{d}_{\mathcal{L}}(f)$.

First, let us examine the function $\text{Upper}_2(x)$ more closely. If $\frac{1}{4} \leq x < \frac{1}{2}$ then $\lfloor \log_2 x \rfloor = -2$ so a simple computation shows that $\text{Upper}_2(x) = 6x^2 - 3x + 1$ in this case. If $\frac{1}{8} \leq x < \frac{1}{4}$ then $\lfloor \log_2 x \rfloor = -3$ so $\text{Upper}_2(x) = 6x^2 + \frac{1}{4}$ in this case. One can check that the function $\text{Upper}_2(x)$ is monotonically increasing on the domain $\left[0, \frac{1}{2}\right]$. (It is continuous, and the derivative exists at all points except those of the form $x = \frac{1}{2^m}$ for some integer $m \geq 1$. The derivative is greater than zero at all points where it exists.) The equation $\text{Upper}_2(x) = \frac{1}{2}$ has only one solution in the interval $\left[0, \frac{1}{2}\right]$, namely $x = \frac{1}{2\sqrt{6}} \in \left[\frac{1}{8}, \frac{1}{4}\right)$. Therefore $\text{Upper}_2(x) \leq \frac{1}{2}$ if and only if $x \leq \frac{1}{2\sqrt{6}}$ (see the first graph in the Appendix).

For the upper bound, from (21) we have

$$\begin{aligned} P_2(f) &\leq \text{Upper}_2\left(\text{d}_{\mathcal{L}}(f)\right), \\ 1 - P_2(f) &= P_2(f+1) \leq \text{Upper}_2\left(\text{d}_{\mathcal{L}}(f+1)\right). \end{aligned}$$

Therefore, using the fact that Upper_2 is monotonic and the assumption $\text{d}_{\mathcal{L}}(f) \leq \text{d}_{\mathcal{L}}(f+1)$ we obtain

$$\begin{aligned} \overline{P}_2(f) &= \min(P_2(f), 1 - P_2(f)) \\ &\leq \min(\text{Upper}_2(\text{d}_{\mathcal{L}}(f)), \text{Upper}_2(\text{d}_{\mathcal{L}}(f+1))) \\ &= \text{Upper}_2(\text{d}_{\mathcal{L}}(f)) = \text{Upper}_2(\text{d}_{\mathcal{A}}(f)). \end{aligned}$$

The bound $\overline{P}_2(f) \leq \frac{1}{2}$ is immediate from $\overline{P}_2(f) = \min(P_2(f), 1 - P_2(f))$.

Now let us deal with the lower bound. If $P_2(f) \leq 1 - P_2(f)$ (in other words, $P_2(f) \leq \frac{1}{2}$), then $\overline{P}_2(f) = P_2(f) \geq \text{Lower}_2(\text{d}_{\mathcal{L}}(f)) = \text{Lower}_2(\text{d}_{\mathcal{A}}(f))$ and we are done. Let us assume $P_2(f) > 1 - P_2(f)$ i.e. $P_2(f) > \frac{1}{2}$. From the behaviour of $\text{Upper}_2(x)$ discussed above, we see that this can only happen when $\text{d}_{\mathcal{L}}(f) \geq \frac{1}{2\sqrt{6}}$. We have to prove that in this case $1 - P_2(f) \geq \text{Lower}_2(\text{d}_{\mathcal{L}}(f))$.

Let us first consider the case $\frac{1}{2\sqrt{6}} \leq \text{d}_{\mathcal{L}}(f) \leq \frac{1}{4}$. We have

$$P_2(f) \leq \text{Upper}_2\left(\text{d}_{\mathcal{L}}(f)\right) = 6\left(\text{d}_{\mathcal{L}}(f)\right)^2 + \frac{1}{4}, \tag{25}$$

therefore

$$\begin{aligned} 1 - P_2(f) &\geq 1 - 6\left(\text{d}_{\mathcal{L}}(f)\right)^2 - \frac{1}{4} = \frac{3}{4} - 6\left(\text{d}_{\mathcal{L}}(f)\right)^2 \\ &\geq 3\text{d}_{\mathcal{L}}(f) - 6\left(\text{d}_{\mathcal{L}}(f)\right)^2 = \text{Lower}_2\left(\text{d}_{\mathcal{L}}(f)\right), \end{aligned}$$

where the last inequality uses the fact that $\text{d}_{\mathcal{L}}(f) \leq \frac{1}{4}$.

Next assume that $\frac{1}{4} < \text{d}_{\mathcal{A}}(f)$. Consider first the subcase $\frac{1}{4} \leq \text{d}_{\mathcal{A}}(f) < \frac{5}{16}$. We have:

$$P_2(f) \leq \text{Upper}_2\left(\text{d}_{\mathcal{L}}(f)\right)$$

and therefore using the fact that $\text{Upper}_2(x) = 6x^2 - 3x + 1$ when $x \geq \frac{1}{4}$ we have

$$1 - P_2(f) \geq 1 - \text{Upper}_2\left(\text{d}_{\mathcal{L}}(f)\right) = 3\text{d}_{\mathcal{L}}(f) - 6\left(\text{d}_{\mathcal{L}}(f)\right)^2 = \text{Lower}_2\left(\text{d}_{\mathcal{L}}(f)\right). \tag{26}$$

Finally, let us consider the subcase $\text{d}_{\mathcal{A}}(f) \geq \frac{5}{16}$. We have

$$1 - P_2(f) = P_2(f+1) \geq \text{Lower}_2\left(\text{d}_{\mathcal{L}}(f+1)\right) \geq \text{Lower}_2\left(\text{d}_{\mathcal{L}}(f)\right), \tag{27}$$

with the last inequality based on the fact that $\frac{5}{16} < \text{d}_{\mathcal{L}}(f) \leq \text{d}_{\mathcal{L}}(f+1)$ and $\text{Lower}_2(x)$ is monotonically increasing when the argument is above $\frac{5}{16}$. □

Remark 12 One might wonder if the situation where $\text{d}_{\mathcal{L}}(f) \leq \text{d}_{\mathcal{L}}(f+1)$ and $P_2(f) > P_2(f+1)$, which is the non-straightforward case in the proof of Theorem 11, does even happen in practice. Experimentally, we did find such functions, but they seemed to be relatively rare. For example, for $n = 6$ and $n = 7$, we generated several random functions for each possible nonlinearity and we only observed that behaviour in a proportion of less than 0.06 of them. Therefore it is a reasonable heuristic, but only a heuristic, to assume that whichever of the functions f and $f+1$ achieves $\min(P_2(f), P_2(f+1))$ also achieves $\min(\text{d}_{\mathcal{L}}(f), \text{d}_{\mathcal{L}}(f+1))$. It also justifies our choice to examine $\min(P_2(f), P_2(f+1))$ for its correlation to $\text{d}_{\mathcal{A}}(f) = \min(\text{d}_{\mathcal{L}}(f), \text{d}_{\mathcal{L}}(f+1))$.

As a byproduct of the proof of Theorem 11, we can also obtain bounds on how large the difference $P_2(f) - P_2(f+1)$ can be when $\text{d}_{\mathcal{L}}(f) \leq \text{d}_{\mathcal{L}}(f+1)$. Namely, denoting $x = \text{d}_{\mathcal{L}}(f)$, we have the following cases. When $x \in \left[0, \frac{1}{2\sqrt{6}}\right]$ we cannot have $P_2(f) > P_2(f+1)$. When $x \in \left(\frac{1}{2\sqrt{6}}, \frac{1}{2}\right]$ both $P_2(f) \leq P_2(f+1)$ and $P_2(f) > P_2(f+1)$ are

possible. If the latter happens, when $x \in \left(\frac{1}{2\sqrt{6}}, \frac{1}{4}\right]$ we obtain from (25) that $P_2(f) - P_2(f+1) = 2P_2(f) - 1 \leq 12x^2 - \frac{1}{2} \leq \frac{1}{4}$; when $x \in \left(\frac{1}{4}, \frac{1}{2}\right]$ we obtain from (26) and (27) that $P_2(f) - P_2(f+1) = 2P_2(f) - 1 \leq 1 - 2\text{Lower}_2(x) \leq \frac{38}{128} \approx 0.296$.

By contrast, for arbitrary functions f, g such that $\text{d}_{\mathcal{L}}(f) < \text{d}_{\mathcal{L}}(g)$ but $P_2(f) > P_2(g)$, the difference $P_2(f) - P_2(g)$ can be larger, approaching 0.5. For example, for any integer $t \geq 2$ consider the functions $f(x) = 1 + x_1x_2 \cdots x_t$ and $g(x) = x_1x_2 + x_3x_4 + \cdots + x_{2t-1}x_{2t}$. Using the calculations in Examples 3, 4, 5 and 9, we have that $\text{d}_{\mathcal{L}}(f) = \frac{1}{2} - \frac{1}{2^t} < \frac{1}{2} - \frac{1}{2^{t+1}} = \text{d}_{\mathcal{L}}(g)$, and $P_2(f) > P_2(g)$ with a difference

$$P_2(f) - P_2(g) = \frac{1}{2} - \frac{3}{2^t} + \frac{13}{2^{2t+1}} \underset{t\to\infty}{\longrightarrow} \frac{1}{2}.$$

An improvement of the lower bound for the BLR linearity test (21) is given in [9]. Namely, it is shown that $P_2(f) \geq H(\text{d}_{\mathcal{L}}(f))$, where

$$H(x) = \begin{cases} 3x - 6x^2 & \text{if } 0 \leq x < \frac{5}{16} \\ \max\left(\frac{45}{128}, \min(g_1(x), g_2(x))\right) & \text{if } \frac{5}{16} \leq x \leq \frac{1}{2}, \end{cases} \tag{28}$$

where for any constant $0 < c \leq \frac{1}{2}$, g_1, g_2 are defined as

$$\begin{aligned} g_1(x) &= x + cx(1-2x)^4, \\ g_2(x) &= x + 2^{12}\left(1 - \frac{5}{4}c + \frac{1}{8}c^2\right)x^3(1-2x)^{12}. \end{aligned}$$

Note that this is indeed an improved lower bound as $\text{Lower}_2(x) \leq H(x)$ and the inequality is strict on the interval $\left(\frac{45}{128}, \frac{1}{2}\right)$. The analogue of Theorem 11 holds for this improved bound as well.

Proposition 13 *With the notations in Theorem* 11*, we have* $\overline{P}_2(f) \geq H(\text{d}_{\mathcal{A}}(f))$*, where* $H(x)$ *is as defined above in (28).*

Proof Examining the proof of Theorem 11 we see that for the lower bound in the interval $\left[\frac{5}{16}, \frac{1}{2}\right]$ the only property that is used is that it is monotonically increasing. It suffices therefore to show that $\min(g_1, g_2)$ is monotonically increasing. We compute g_1' and g_2', the derivatives of g_1 and g_2, and show that they are positive on the specified domain. Namely, $0 < c \leq \frac{1}{2}$ implies $0 < 1 - \frac{5}{4}c + \frac{1}{8}c^2 \leq 1$; further, $\frac{5}{16} \leq x \leq \frac{1}{2}$ implies $1 - 2x \leq \frac{3}{8}$, $10x - 1 \leq 4$ and $x(1-2x) \leq \frac{15}{2^7}$. Therefore,

$$\begin{aligned} g_1'(x) &= 1 - c(1-2x)^3(10x-1) \geq 1 - \frac{1}{2} \cdot \frac{3^3}{8^3} \cdot 4 = 1 - \frac{3^3}{2^8} > 0, \\ g_2'(x) &= 1 - 3 \cdot 2^{12}\left(1 - \frac{5}{4}c + \frac{1}{8}c^2\right)x^2(1-2x)^{11}(10x-1) \\ &\geq 1 - 3 \cdot 2^{12} \cdot \frac{15^2}{2^{14}} \cdot \frac{3^9}{8^9} \cdot 4 = 1 - \frac{3^{12} \cdot 5^2}{2^{27}} > 0. \end{aligned}$$

This concludes the proof. □

6 Estimating nonlinearity

The above affinity tests can be used to estimate the nonlinearity of a Boolean function. The probability of failing a test can be estimated by running the test several times and using statistical methods such as the binomial proportion confidence interval (see [14]). The bounds will then allow to give an interval for the value of the nonlinearity as per (1) and (2). For simplicity, we will assume that we have obtained an exact value for $P_k(f)$ (in practice we will actually obtain a confidence interval). We will examine each test in turn. The graphs in the Appendix will aid the discussion.

We first look at the affine test based on the BLR test, as described in Section 5. The first graph in the Appendix displays the lower and upper bound described in Theorem 11. Thus, for values of $0 \le \overline{P}_2(f) < y_1^{(2)} = \frac{45}{128} = 0.3515625$ we can estimate the nonlinearity with good precision as being in the interval $\mathrm{d}_{\mathcal{A}}(f) \in \left[\mathrm{Upper}_2{}^{-1}(\overline{P}_2(f)), \mathrm{Lower}_2{}^{-1}(\overline{P}_2(f))\right]$. The length of this interval increases with $\overline{P}_2(f)$ to a length of approximately 0.058. For $\overline{P}_2(f) = \frac{45}{128}$ we get

$$\mathrm{d}_{\mathcal{A}}(f) \in \left[\mathrm{Upper}_2{}^{-1}\left(\frac{45}{128}\right), \frac{3}{16}\right] \cup \left[\frac{5}{16}, \frac{45}{128}\right].$$

For $\frac{45}{128} < \overline{P}_2(f) < \frac{1}{4}$, $\mathrm{Lower}_2{}^{-1}(\overline{P}_2(f)) = \{\alpha^{(2)}(\overline{P}_2(f)), \beta^{(2)}(\overline{P}_2(f)), \overline{P}_2(f)\}$, where $0 < \alpha^{(2)}(y) \le \beta^{(2)}(y) < \frac{5}{16}$ are the two roots of the equation $3x - 6x^2 = y$ in this domain. We obtain two disjoint intervals where $\mathrm{d}_{\mathcal{A}}(f)$ might be:

$$\mathrm{d}_{\mathcal{A}}(f) \in \left[\mathrm{Upper}_2{}^{-1}(\overline{P}_2(f)), \alpha^{(2)}(\overline{P}_2(f))\right] \cup \left[\beta^{(2)}(\overline{P}_2(f)), \overline{P}_2(f)\right].$$

Finally, for $\overline{P}_2(f) \ge \frac{1}{4}$, the interval for $\mathrm{d}_{\mathcal{A}}(f)$ is $\left[\mathrm{Upper}_2{}^{-1}(\overline{P}_2(f)), \overline{P}_2(f)\right]$. The estimate for $\mathrm{d}_{\mathcal{A}}(f)$ becomes less and less precise (the interval length increases) as $\overline{P}_2(f)$ increases. When $\overline{P}_2(f)$ reaches $\frac{1}{2}$, we obtain $\mathrm{d}_{\mathcal{A}}(f) \in \left[\frac{1}{2\sqrt{6}}, \frac{1}{2}\right)$, an interval of length approximately 0.295.

Next we look at the nonhomomorphicity test $f(u_1 + \cdots + u_k) + f(u_1) + \cdots + f(u_k) = 0$ with odd $k \ge 3$. We use the upper bound Upper_k described in Theorem 2 and the lower bound Lower_k described in Corollary 8, illustrated for $k = 3, 5$ in the second and third graph in the Appendix.

As x increases in the interval $[0, 0.5]$, $\mathrm{Upper}_k(x)$ increases, whereas $\mathrm{Lower}_k(x)$ first increases from 0 to a local maximum $y_2^{(k)}$, then decreases to a value of $y_1^{(k)} = \frac{1}{2}\left(1 - \frac{1}{2^{k-1}}\right)$ (reached for $x = \frac{1}{4}$) and increases again to 0.5. Consequently, we have three cases. When $0 \le P_k(f) < y_1^{(k)}$ we have that

$$\mathrm{d}_{\mathcal{A}}(f) \in \left[\frac{1}{2}\left(1 - \sqrt[k+1]{1 - 2P_k(f)}\right), \alpha^{(k)}(P_k(f))\right],$$

where for each $0 \le y \le 0.5$ we denote by $0 < \alpha^{(k)}(y) \le \beta^{(k)}(y) < \frac{1}{2}$ the two roots of the equation $\frac{1}{2}\left(1 - (1-2x)^{k+1} - 2^{k+1}x^k(1-x)\right) = y$. The length of this interval increases

as $P_k(f)$ increases from 0 to $y_1^{(k)}$ (for illustration, it increases to a value of 0.028, 0.016 and 0.011 for $k = 3, 5$ and 7, respectively).

When $y = P_k(f) \in [y_1^{(k)}, y_2^{(k)}]$, we have that

$$\mathrm{d}_{\mathcal{A}}(f) \in \left[\frac{1}{2}\left(1 - \sqrt[k+1]{1-2y}\right), \alpha^{(k)}(y)\right] \cup \left[\beta^{(k)}(y), \frac{1}{2}\left(1 - \sqrt[k-1]{1-2y}\right)\right].$$

Finally, for $y_2^{(k)} < P_k(f) \leq 0.5$, we have

$$\mathrm{d}_{\mathcal{A}}(f) \in \left[\frac{1}{2}\left(1 - \sqrt[k+1]{1-2P_k(f)}\right), \frac{1}{2}\left(1 - \sqrt[k-1]{1-2P_k(f)}\right)\right]. \tag{29}$$

Note that the less tight bounds (7) from [14] would give the considerably less accurate estimate

$$\mathrm{d}_{\mathcal{A}}(f) \in \left[0, \frac{1}{2}\left(1 - \sqrt[k+1]{\frac{1-2P_k(f)}{2^n}}\right)\right].$$

The length of the interval produced by our estimate (29) is

$$\frac{1}{2}\left(\sqrt[k+1]{1-2P_k(f)} - \sqrt[k-1]{1-2P_k(f)}\right).$$

This quantity has a unimodal behavior: the length increases as a function of $P_k(f)$, peaking at a value of

$$\frac{1}{2}\left(\left(\frac{k-1}{k+1}\right)^{\frac{k-1}{2}} - \left(\frac{k-1}{k+1}\right)^{\frac{k+1}{2}}\right),$$

achieved when

$$P_k(f) = \frac{1}{2}\left(1 - \left(\frac{k-1}{k+1}\right)^{\frac{k^2-1}{2}}\right),$$

and then decreases to 0, when $P_k(f)$ reaches 0.5. For example, if $k = 3, 5, 7$, the length of the interval peaks at a value of 0.125, 0.0741 and 0.05273, respectively (achieved when $P_k(f) = 0.469, 0.496$ and 0.4995, respectively). The maximum length of the interval is achieved when $P_k(f)$ is quite close to 0.5; the larger the value of k, the smaller the maximum length of the interval, that is, the more precisely we can estimate the nonlinearity.

We summarize these results in Table 1, which contains, for different values of k, the maximum length of the interval obtained when estimating the nonlinearity. The length is

Table 1 Precision of estimating the nonlinearity

k	Length of interval for $\mathrm{d}_{\mathcal{A}}$ when P_k is low	Length of interval for $\mathrm{d}_{\mathcal{A}}$ when P_k is high
2	≤ 0.058	≤ 0.295
	$\overline{P}_2 \leq 0.3515625$	$0.3515625 \leq \overline{P}_2 \leq 0.5$
3	≤ 0.028	≤ 0.125
	$P_3 \leq 0.375$	$0.375 \leq P_3 \leq 0.5$
5	≤ 0.016	≤ 0.0741
	$P_5 \leq 0.46875$	$0.46875 \leq P_5 \leq 0.5$
7	≤ 0.011	≤ 0.05273
	$P_7 \leq 0.492188$	$0.492188 \leq P_7 \leq 0.5$

Table 2 Examples of estimating the nonlinearity

Function	$\mathrm{d}_{\mathcal{A}}$	$k=3$, estimated $\mathrm{d}_{\mathcal{A}}$	$k=5$, estimated $\mathrm{d}_{\mathcal{A}}$	$k=7$, estimated $\mathrm{d}_{\mathcal{A}}$
x_1x_2	$\frac{1}{4}$	$\left[\frac{1}{4}-0.10355,\ \frac{1}{4}\right]$	$\left[\frac{1}{4}-0.06498,\ \frac{1}{4}\right]$	$\left[\frac{1}{4}-0.0473,\ \frac{1}{4}\right]$
$x_1x_2+x_3x_4$	$\frac{3}{8}$	$\left[\frac{3}{8}-0.125,\ \frac{3}{8}\right]$	$\left[\frac{3}{8}-0.07343,\ \frac{3}{8}\right]$	$\left[\frac{3}{8}-0.05178,\ \frac{3}{8}\right]$
$x_1x_2+x_3x_4$ $+x_5x_6$	$\frac{7}{16}$	$\left[\frac{7}{16}-0.11428,\ \frac{7}{16}\right]$	$\left[\frac{7}{16}-0.06250,\ \frac{7}{16}\right]$	$\left[\frac{7}{16}-0.04261,\ \frac{7}{16}\right]$
$x_1x_2+x_3x_4$ $+x_5x_6+x_7x_8$	$\frac{31}{32}$	$\left[\frac{31}{32}-0.09375,\ \frac{31}{32}\right]$	$\left[\frac{31}{32}-0.04750,\ \frac{31}{32}\right]$	$\left[\frac{31}{32}-0.03125,\ \frac{31}{32}\right]$
$x_1x_2x_3$	$\frac{1}{8}$	$\left[\frac{1}{8}-7.85\cdot10^{-3},\ \frac{1}{8}\right]$	$\left[\frac{1}{8}-5.98\cdot10^{-4},\ \frac{1}{8}\right]$	$\left[\frac{1}{8}-5.00\cdot10^{-5},\ \frac{1}{8}\right]$
$x_1x_2x_3x_4$	$\frac{1}{16}$	$\left[\frac{1}{16}-6.82\cdot10^{-4},\ \frac{1}{16}\right]$	$\left[\frac{1}{16}-9.30\cdot10^{-6},\ \frac{1}{16}\right]$	$\left[\frac{1}{16}-1.42\cdot10^{-7},\ \frac{1}{16}\right]$
$x_1\cdots x_5$	$\frac{1}{32}$	$\left[\frac{1}{32}-7.17\cdot10^{-5},\ \frac{1}{32}\right]$	$\left[\frac{1}{32}-2.13\cdot10^{-7},\ \frac{1}{32}\right]$	$\left[\frac{1}{32}-7.09\cdot10^{-10},\ \frac{1}{32}\right]$
$x_1\cdots x_6$	$\frac{1}{64}$	$\left[\frac{1}{64}-8.26\cdot10^{-6},\ \frac{1}{64}\right]$	$\left[\frac{1}{64}-5.73\cdot10^{-9},\ \frac{1}{64}\right]$	$\left[\frac{1}{64}-4.47\cdot10^{-12},\ \frac{1}{64}\right]$

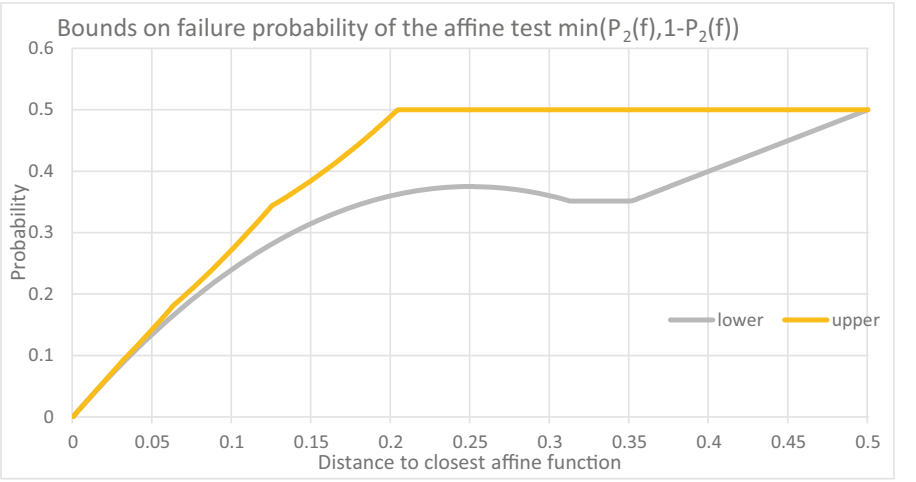
displayed firstly, for low values of $P_k(f)$, namely the values in the interval $[0, y_1^{(k)}]$ discussed above. Secondly, the last column displays the maximum length of the interval for the remaining (higher) values of $P_k(f)$.

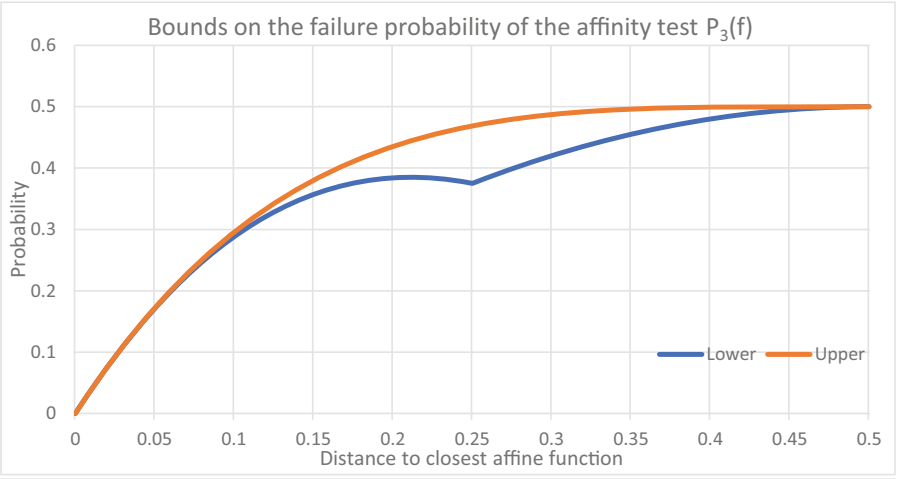
We also present in Table 2 the estimate of the nonlinearity $\mathrm{d}_{\mathcal{A}}$ that would be obtained by this method for a few examples of functions, and compare it with the true value of the nonlinearity. The examples in this table are the ones in Example 5, $f(x_1,\ldots,x_n) = x_1x_2+\cdots+x_{m-1}x_m$ with $m=2,4,6,8$ and $n\geq m$ and the functions in Example 9 of the type $f(x_1,\ldots,x_n)=x_1x_2\cdots x_m$ with $m=3,4,5,6$ and $n\geq m$. We observe that for all these functions, the true value of the nonlinearity is at the top end of the estimated interval.

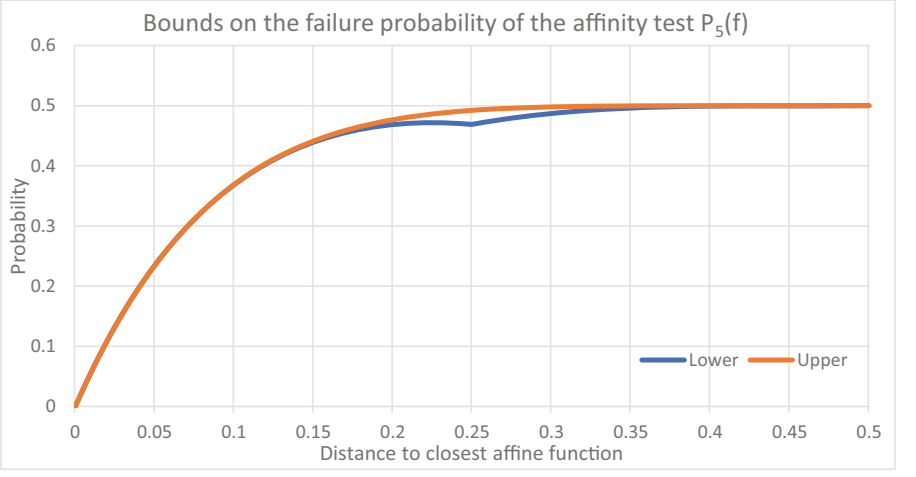
We also examined experimentally random functions in up to 9 variables (see the fourth figure in the Appendix), plotting the probability of failure $P_3(f)$ as a function of the nonlinearity $\mathrm{d}_{\mathcal{A}}(f)$. To obtain data for each possible value of the nonlinearity we started by randomly generating several functions for each possible weight lower than 0.5. We then computed their nonlinearity (for weights lower than 0.25 it is equal to the weight of the function, as the function is closer to the all-zero function than to any other affine function; for higher weights, the nonlinearity can be different from the weight, but many functions will have a nonlinearity close to their weight). We noticed that for functions in 7 or more variables most of the functions in our data have probability P_3 of failing the test close to the upper bound for P_3. This translates to the true value of the nonlinearity being at the low end of the estimated interval. We observed a similar situation for $k=5,7$.

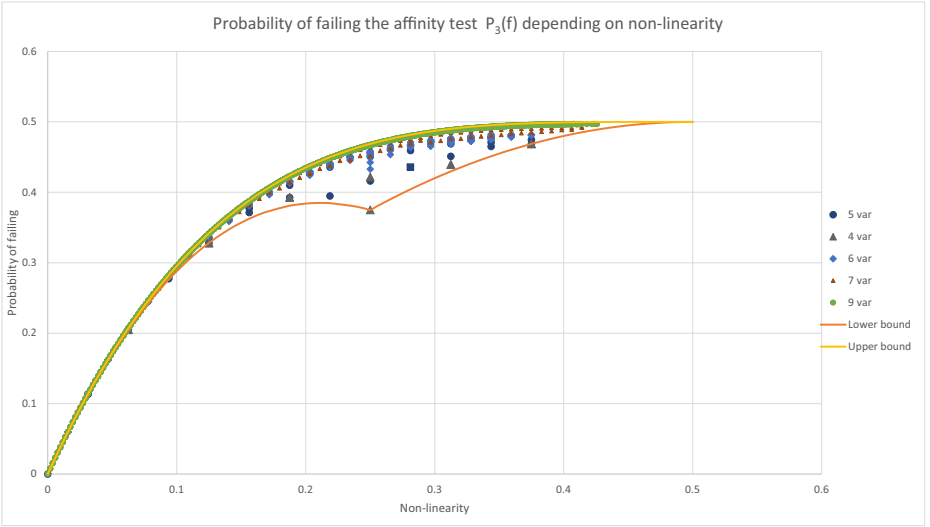
To conclude this section, we note that each test we considered is quite accurate in estimating nonlinearity when the probability of failing the test is small (and consequently the nonlinearity of the function is small), but the accuracy decreases as the probability of failing the test increases. If we were to apply different tests to the same function, we note that the estimated interval for the nonlinearity is least accurate when using the affinity test based on the BLR test. The tests based on $(k+1)$-st order nonhomomorphicity with k odd have better accuracy, and this accuracy improves as k increases.

Appendix


Bounds on failure probability of the affine test min(P2(f),1-P2(f))
Probability
Distance to closest affine function
lower
upper


Bounds on the failure probability of the affinity test P3(f)
Probability
Distance to closest affine function
Lower
Upper


Bounds on the failure probability of the affinity test P5(f)
Probability
Distance to closest affine function
Lower
Upper



Acknowledgements The authors are grateful to the editor for the prompt handling of our paper and to the reviewers for extensive and helpful comments and suggestions which have highly improved the manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Bellare, M., Coppersmith, D., Håstad, J., Kiwi, M., Sudan, M.: Linearity testing in characteristic two. IEEE Trans. Inf. Theory **42**(6), 1781–1795 (1996)
2. Bera, D., Maitra, S., Roy, D., Stănică, P.: Limitation of the BLR testing in estimating nonlinearity. In: Workshop on Coding and Cryptography, Rennes, France, Paper #50 (2019)
3. Blum, M., Luby, M., Rubinfeld, R.: Self-testing/correcting with applications to numerical problems. J. Comput. Syst Sci. **47**(3), 549–595 (1993)
4. Bullen, P.S.: Handbook of means and their inequalities. Springer (2003)
5. Cusick, T.W., Stănică, P. Cryptographic Boolean Functions and Applications, 2nd edn. Academic Press, San Diego (2017)
6. Dinur, I., Shamir, A.: Cube attacks on tweakable black box polynomials. Adv. in Crypt. – EUROCRYPT pp. 278–299, LNCS 5479. Springer, Berlin (2009)
7. Dinur, I., Shamir, A.: Applying cube attacks to stream ciphers in realistic scenarios. Cryptogr. Communic. **4**(3-4), 217–232 (2012)
8. Doğanaksoy, A., Sağdiçoğlu, S., Saygi, Z., Uğuz, M.: A note on linearity and homomorphicity. In: Michon, J.-F., Valarcher, P., Yunès, J.-B. (eds.) Boolean Functions: Cryptography and Applications, pp. 280–295 (2006)
9. Kaufman, T., Litsyn, S., Xie, N.: Breaking the ϵ-soundness bound of the linearity test over $GF(2)$. SIAM J. Computing **39**(5), 1988–2003 (2010)

10. Molland, H., Helleseth, T.: An Improved Correlation Attack Against Irregular Clocked and Filtered Keystream Generators. Adv. in Crypt. – CRYPTO 2004, pp. 373–389. Springer, Berlin (2004)
11. Sălăgean, A., Stănică, P.: Estimating the nonlinearity of Boolean functions using probabilistic linearity tests. Proc. Sequences and Their Applications – SETA 2020, Paper #45 (2020)
12. Vielhaber, M., Breaking, O.N.E.: FIVIUM by AIDA an algebraic IV differential attack. Cryptology ePrint Archive Report 2007/413. http://eprint.iacr.org/ (2007)
13. Winter, R., Sălăgean, A., Phan, R.C.W.: Comparison of cube attacks over different vector spaces. In: Groth, J. (ed.) 15th IMA International Conference on Cryptography and Coding, IMACC, LNCS 9496, pp. 225–238. Springer (2015)
14. Zhang, X.-M., Zheng, Y.: The nonhomomorphicity of Boolean functions. In: Tavares, S., Meijer, H. (eds.) Selected Areas in Cryptography, SAC, pp. 280–295. Springer (1999)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.