Publication II

Miia Hermelin and Kaisa Nyberg. 2008. Multidimensional linear distinguishing attacks and Boolean functions. In: Jean-Francis Michon, Pierre Valarcher, and Jean-Baptiste Yunès (editors). Preproceedings of the Fourth International Workshop on Boolean Functions: Cryptography and Applications (BFCA 2008). Copenhagen, Denmark. 19-21 May 2008. Publications des Universités de Rouen et du Havre.

# MULTIDIMENSIONAL LINEAR DISTINGUISHING ATTACKS AND BOOLEAN FUNCTIONS

Miia Hermelin[1] and Kaisa Nyberg[1]

**Abstract**. In this paper theoretical aspects of multidimensional linear distinguishing attacks are investigated. Using known examples of highly nonlinear Boolean functions we demonstrate how multidimensional linear approximations offer significant reduction in data complexity in distinguishing attacks. We also get concrete examples where one-dimensional linear approximations are never statistically independent.

## 1. Introduction

Linear cryptanalysis method was introduced by Matsui in [1] where two statistical key-recovery attacks on the DES, Algorithm 1 and Algorithm 2, were presented. Later linear approximations have also been used for distinguishing an output sequence of a key stream generator from a truly random sequence.

Enhancements of the linear cryptanalysis method using multiple linear approximations were presented by Kaliski and Robshaw in [2] and Biryukov, et al., in [3]. Truly multidimensional linear distinguishing attack was presented by Englund and Maximov in [4], and the foundations of the statistical analysis were presented by Baignères, et al., in [5].

The goal of this paper is to investigate theoretical aspects of distinguishing attacks based on multidimensional linear approximation. For this purpose, we interpret a linear approximation as a Boolean function, and show that its strength can be measured using the $\ell_2$-distance between the related probability distribution and the uniform distribution.

[1] Helsinki University of Technology email: miia.hermelin@tkk.fi

We call this measure capacity due to its similarity to the notion of capacity used by Biryukov, et al., in [3]. We show that the probability distribution of a multidimensional linear distinguisher can be determined based on the one-dimensional linear approximations. This approach allows us to select the strongest linear approximations and is computationally more flexible compared to the approach of handling the full probability distributions in [4].

We also explore the limits of multidimensional linear distinguishing attacks. We show that the vector bent functions offer the best resistance, but also that the gain of multidimensional approximations compared to one-dimensional approximations is the largest for bent functions. We investigate the probability distribution of the basic multidimensional linear distinguisher of the filter generator, see [6], and calculate its capacity for some highly nonlinear filter functions. In all these examples, we can observe significant reduction in data complexity compared to the one-dimensional linear approximation. We also see concrete examples where the assumption of statistical independence of linear approximations, which is the basis of the theory in [2] and [3], does not hold.

Finally, we investigate how multidimensional linear approximations can be chained for composition of independent cipher layers and how the probability distribution of the chained approximation can be determined. We present the multidimensional form of the Piling-up lemma and prove an upper bound to the capacity of chained or parallel approximations.

The outline of this paper is as follows. In Section 2 the basic concepts of vector Boolean functions and related probability distributions are introduced. The theory of multidimensional linear distinguishing attacks is presented in Section 3. The properties of highly nonlinear Boolean functions are studied in Section 4, and they are examined further in the context of a filter key stream generator in Section 5. The results on compositions of multidimensional approximations are presented in Section 6, and we conclude in Section 7.

## 2. Probability Distribution of a Boolean Function

We will denote the space of $n$-dimensional binary vectors by $V_n$. The inner product is defined for $a = (a_1, \ldots, a_n), b = (b_1, \ldots, b_n) \in V_n$ as $a \cdot b = a_1 b_1 + \cdots + a_n b_n$, where + is sum modulo 2. If necessary, $\bigoplus_{i=1}^{m} a_i$ is used to notate the sum $a_1 + \cdots + a_m$ modulo 2.

A function $f : V_n \to V_1$ is called a Boolean function. A function $f : V_n \to V_m$, $f = (f_1, \ldots, f_m)$, where $f_i$ are Boolean functions is called a vector Boolean function of dimension $m$. A linear Boolean function from $V_n \to V_m$ is represented by an $m \times n$ binary matrix $W$. The $m$ rows of $W$ are denoted by $w_1, \ldots, w_m$, where each $w_i$ is a binary vector of length $n$.

Let $f, g : V_n \to V$ be Boolean functions. The correlation between $f$ and $g$ is $c(f, g) = 2^{-n}\left(\#\{\xi \in V_n \mid f(\xi) = g(\xi)\} - \#\{\xi \in V_n \mid f(\xi) \neq g(\xi)\}\right)$. If $g = 0$, then $c(f, 0) = c(f)$ is called the correlation of $f$.

Let $f : V_n \to V_m$. We will call the vector $p(f) = (p_0(f), \ldots, p_{2^m-1}(f))$, where $p_\eta(f) = 2^{-n}\#\{\xi \in V_n | f(\xi) = \eta\}$, the probability distribution (p.d.) of $f$. We may also denote $p_\eta(f)$ by $p_\eta$ if the function $f$ is clear from the context. The vector $\theta_m = 2^{-m}(1, \ldots, 1) \in \mathbb{R}^{2^m}$ is used to denote the $2^m$-valued uniform p.d.

Let $\phi : V_n \to \mathbb{R}$ be a real-valued function. The Walsh-Hadamard transform $\hat{\phi}$ of $\phi$ is defined as

$$\hat{\phi}(u) = \sum_{\xi \in V_n} \phi(\xi)(-1)^{\xi \cdot u},\ u \in V_n.$$

Then $\phi(\xi) = 2^{-n}\hat{\hat{\phi}}(\xi), \xi \in V_n$, using the inverse of Walsh-Hadamard transform. The convolution of two functions $\phi : V_n \to \mathbb{R}$ and $\psi : V_n \to \mathbb{R}$ is defined as

$$(\phi * \psi)(\eta) = \sum_{\xi \in V_n} \phi(\xi)\psi(\xi + \eta),\ \eta \in V_n.$$

It is straightforward to verify that then

$$\widehat{(\phi * \psi)}(u) = \hat{\phi}(u)\hat{\psi}(u),\ u \in V_n. \tag{1}$$

If $\phi(\xi) = (-1)^{f(\xi)}$ for a Boolean function $f : V_n \to V$, then $\hat{\phi}(u)$ is denoted by $\hat{f}(u)$. The set $\{\hat{f}(u) \mid u \in V_m\}$ is called the Walsh spectrum of $f$. Parseval's theorem states that $2^n \sum_\xi \phi(\xi)^2 = \sum_u \hat{\phi}(u)^2$. For a Boolean function $f : V_n \to V$ it then follows that

$$2^{-2n} \sum_{u \in V_n} \hat{f}(u)^2 = \sum_{u \in V_n} c(f(\xi), u \cdot \xi)^2 = 1. \tag{2}$$

For $a \in V_m$, we use notation $\rho(a) = c(a \cdot f)$. Then we have

$$\rho(a) = 2^{-n} \sum_{\xi \in V_n} (-1)^{a \cdot f(\xi)} = \sum_{\eta \in V_m} (-1)^{a \cdot \eta} p_\eta = \hat{p}(a). \tag{3}$$

Using the inverse Walsh-Hadamard transform we get the following lemma.

**Lemma 2.1.** *Suppose that $f : V_n \to V_m$ is a Boolean function with p.d. $p$ and with one-dimensional correlations $\rho(a)$ of $a \cdot f$. Then*

$$p_\eta = 2^{-m} \sum_{a \in V_m} (-1)^{a \cdot \eta} \rho(a), \text{ for all } \eta \in V_m.$$

## 3. Multidimensional Linear Distinguishing Attacks

Linear cryptanalysis can be significantly enhanced by using multiple linear approximations as shown most notably in [2] and [3]. However, these approaches are restricted by the assumption that the individual linear approximations are statistically independent. This assumption was also studied by S. Murphy in [7]. Baignères, et al., developed the statistical theory of general multidimensional distinguishing [5]. Recently, we presented in [8] a truly multidimensional generalisation of Matsui's Algorithm 1, which does not assume statistical independence of the linear approximations and also performs in practice better than the algorithms of [2] and [3]. In this section, we will give the theoretical foundations of the multidimensional linear distinguishing attack. First we will define a multidimensional linear approximation as a vector Boolean function and then consider the statistical properties of the p.d. of this Boolean function.

### 3.1. Multidimensional Approximation of Boolean Functions

Let $f : V_n \to V_l$ be a vector Boolean function and binary vectors $w_i \in V_l$ and $u_i \in V_n$, $i = 1, 2, \ldots, , m$, be linear masks such that the paired masks $(u_i, w_i)$ are linearly independent. Let us define functions $g_i$ by

$$g_i(\xi) \coloneqq w_i \cdot f(\xi) + u_i \cdot \xi, \tag{4}$$

and denote their correlations by $\rho_i = c(g_i)$, $i = 1, 2, \ldots, m$. We will call these correlations the base correlations, and the corresponding linear approximations of $f$ the base approximations. We investigate the p.d. of the $m$-dimensional vector Boolean function $g(\xi) \coloneqq W f(\xi) + U\xi$, where $W = (w_1, \ldots, w_m)$, $U = (u_1, \ldots, u_m)$ and $g = (g_1, \ldots, g_m)$. Let the p.d. of $g$ be $p$ and that the components $a \cdot g$ have correlations $\rho(a) = c(a \cdot g)$, $a \in V_m$. If $e_i = (0 \ldots 010 \ldots 0)$ with 1 at the $i$th place, then $\rho(e_i) = \rho_i, i = 1, \ldots, m$. Given the one-dimensional correlations $\rho(a) = c(a \cdot g)$, $a \in V_m$, the probability distribution $p(g)$ can be determined using Lemma 2.1.

### 3.2. **Capacity of a Multidimensional Linear Distinguisher**

The strength of a linear multidimensional approximation is determined by the nonuniformity of its p.d., which is measured using its capacity to be defined next.

**Definition 3.1.** Let $p = (p_0, \ldots, p_M)$ and $q = (q_0, \ldots, q_M)$ be two p.d.'s. Their (mutual) capacity is

$$C(p, q) = \sum_{\eta=0}^{M} \frac{(p_\eta - q_\eta)^2}{q_\eta}.$$

If $M = 2^m - 1$ and $q = \theta_m$ is uniform then $C(p, \theta_m) = 2^m \|p - \theta_m\|_2^2$ will be called the capacity of $p$ and we will denote it by $C(p)$. It is also called the Squared Euclidean Imbalance [5]. If $p$ is the p.d. of a Boolean function $g$, then we set $C(p) = C(g)$ and call $C(g)$ the capacity of $g$.

The following corollary of Lemma 2.1 is obtained using Parseval's theorem. An equivalent form of it can be found in [5], where the proof was based on the inverse Walsh-Hadamard transform of the multidimensional biases $p_\eta(g) - 2^{-m}$.

**Corollary 3.2.** *Let $g$ be a Boolean function with p.d. $p$. Then*

$$C(g) = \sum_{a \neq 0} \rho(a)^2.$$

A distinguishing attack can be described as a hypothesis testing problem. Null hypothesis $H_0$ states that the empirical data $\mathbf{z}^N$ of $N$ words is derived from p.d. $p$ and the alternative hypothesis $H_1$ states that $\mathbf{z}^N$ is derived from p.d. $q \neq p$ (see [9]).

The following theorem was proved in [5] where the log-likelihood-ratio was used as the distinguisher of the multidimensional hypothesis testing problem. Note that the theorem makes no assumptions about statistical independence of the base approximations.

**Theorem 3.3.** *Assume that the p.d.'s $p$ and $q$ are close to each other: $|q_\eta - p_\eta| \ll q_\eta$, for all $\eta \in V_m$. Then the amount of data needed to distinguish between $H_0$ and $H_1$ is*

$$N = \frac{\gamma}{C(p, q)}, \tag{5}$$

*where $\gamma$ depends on the level and the power of the test.*

Hence, if we want to distinguish whether $\mathbf{z}^N$ comes from a cipher with p.d. $p$ or from a random source with uniform p.d., the amount of data needed using $m$ linear approximations is

$$N_m = \frac{\gamma}{C(p)} = \frac{\gamma}{\sum_{a\neq 0} \rho(a)^2}, \tag{6}$$

where $\gamma$ depends on the level and the power of the test. In the one-dimensional case, we have a linear approximation such as (4). Let $\rho$ be the correlation of the approximation. The number of bits $N_1$ needed to distinguish $\mathbf{z}^N$ from a random sequence is $\gamma/\rho^2$.

We can see that multidimensional approximation offers significant reduction in data complexity, in particular, for functions with one-dimensional linear approximations with uniformly small correlations. In Section 5, we will see examples where all the correlations $\rho(a)$ are equal or their absolute values are equal. Then $N_m = N_1/(2^m - 1)$.

## 4. Optimal and Near Optimal Capacity of Boolean Functions

In this section, we determine the capacities of some known examples of highly nonlinear vector Boolean functions, which offer strong resistance against basic one-dimensional linear attacks. First we will show that vector bent functions are optimal against multidimensional linear cryptanalysis and determine its capacity.

### 4.1. Bent Functions

Multidimensional Walsh transform was introduced in [10].

**Definition 4.1.** Let $f : V_n \to V_m$ be a Boolean function. Then we define

$$\mathcal{W}_{f+U}(x) = 2^{-n} \sum_{\xi\in V_n} \prod_{i=1}^{m} x_i^{f_i(\xi)\oplus u_i\cdot\xi},$$

where the sum is taken in the set $\mathbb{Z}[x_1, \ldots, x_m]/\langle x_1^2 - 1, \ldots, x_m^2 - 1\rangle$ of multivariate polynomials over integers where the indeterminates $x_i$ satisfy $x_i^2 = 1$. The transform that maps $f$ to the mapping $U \to \mathcal{W}_{f+U}(x), U = (u_1, \ldots, u_m) \in V_n^m$, is called the multi-Walsh transform.

Let the p.d. of $f$ be $p$. If $U = 0$, we can give the transform as follows:

$$\mathcal{W}_f(x) = \sum_{\eta \in V_m} p_\eta \prod_{i=1}^{m} x_i^{\eta_i}. \tag{7}$$

The polynomial $\Theta_m(x) = 2^{-m} \sum_{\eta \in V_m} \prod_i x_i^{\eta_i}$ corresponds to the uniform distribution $\theta_m$. The following theorem is the multidimensional equivalent of Parseval's theorem, and its proof can be found in [10].

**Theorem 4.2.** *For any vector-valued Boolean function $f : V_n \to V_m$ the following holds:*

$$\sum_{U \in V_n^m} \mathcal{W}_{f+U}^2(x) = 2^{(m-1)n} \left(1 + (2^n - 1)\Theta_m(x)\right).$$

In the one-dimensional case, bent functions are defined as functions which have equally small correlations, in absolute value, to all linear functions, see (2). Analogically, the multi-bent functions are defined as functions with uniform multi-Walsh spectrum as follows.

**Definition 4.3.** A vector valued Boolean function $f : V_n \to V_m$ is multi-bent if

$$\mathcal{W}_{f+U}^2(x) = 2^{-n}(1 + (2^n - 1)\Theta_m(x))$$

for all $U \in V_n^m$.

It was shown in [10] that $f$ is multi-bent if and only if it is bent in the classical sense (i.e., its components $a \cdot f(x)$, $a \neq 0$, are bent). Hence, $f$ is multi-bent if and only if $W \circ f \circ T + U$ is multi-bent for all linear transformations $W$ and $U$ and linear bijections $T$. We have the following theorem considering the capacity of bent functions.

**Theorem 4.4.** *The capacity of a multi-bent Boolean function $f : V_n \to V_m$ satisfies*

$$C(f) = 2^{m-n} - 2^{-n}. \tag{8}$$

*Proof.* By (7), it is straightforward to verify that the polynomial $\mathcal{W}_f^2(x)$ corresponds to the p.d. $p * p$. The constant term in the polynomial is $\sum_\eta p_\eta^2 = 2^{-n}(1 + (2^n - 1)2^{-m})$. By Definition 3.1 $C(f) = 2^m \|p - \theta_m\|_2^2 = 2^m \sum_\eta p_\eta^2 - 1$, from where the claim follows. □

If $f$ is multi-bent, then $f + U$ is multi-bent and $C(f + U) = 2^{m-n} - 2^{-n}$, for all $U : V_n \to V_m$ linear. It follows that multi-bent functions are optimal against multidimensional linear cryptanalysis. It is an open question whether there are functions other than multi-bent that satisfy (8).

### 4.2. **Power Functions**

We will next study functions $f : V_n \to V_n$ that are of the form $f(x) = x^d$, $d \geq 1$. We denote the approximations by $g(x) = Wf(x) + Ux$, and restrict to cases where $W$ is invertible and $Ux = ux$, for some $u \in \mathbb{F}_{2^n}$.

The trace is denoted by $\mathrm{Tr}(\alpha) = \alpha + \alpha^2 + \cdots + \alpha^{2^{n-1}}$, for $\alpha \in \mathbb{F}_{2^n}$. For all $a \in V_n$, there exists a unique $\alpha \in V_n$ such that $a \cdot x = \mathrm{Tr}(\alpha x)$. Now we can replace $\rho(a)$ in Corollary 3.2 with $\rho'(\alpha) = \sum_{x \in V_n} (-1)^{\mathrm{Tr}(\alpha g(x))}$. Similarly, we introduce the following notation

$$\mathcal{W}'_h(\alpha) = \sum_{x \in V_n} (-1)^{h(x)+\mathrm{Tr}(\alpha x)} = \sum_{x \in V_n} (-1)^{h(x)+a \cdot x} = \mathcal{W}_h(a),$$

where $h : V_n \to V$ is a Boolean function. We can now state the following lemma that gives a connection between the modified correlation and Walsh transform of a power function:

**Lemma 4.5.** *Let $f(x) = x^d$, $d \geq 1$ and $g(x) = f(x) + ux$, $u \in V_m$. Then*

$$\rho'(\alpha^d) = \mathcal{W}'_{\mathrm{Tr}(f)}(\alpha^{d-1}u). \tag{9}$$

*Proof.* We use the definitions of $\rho'(\alpha)$ and $\mathcal{W}'_h(\alpha)$ to get

$$\begin{aligned} \rho'(\alpha^d) &= \sum_{x \in V_n} (-1)^{\mathrm{Tr}(\alpha^d g(x))} = \sum_{x \in V_n} (-1)^{\mathrm{Tr}((\alpha x)^d)+\mathrm{Tr}(\alpha^d u x)} \\ &= \sum_{y \in V_n} (-1)^{\mathrm{Tr}(y^d)+\mathrm{Tr}(\alpha^{d-1} u y)} = \mathcal{W}'_{\mathrm{Tr}(f)}(\alpha^{d-1}u). \end{aligned}$$

□

**Theorem 4.6.** *If $\gcd(d, 2^n - 1) = \gcd(d - 1, 2^n - 1) = 1$ then the linear approximation $g(x) = x^d + ux$, $u \in V_n$ has capacity $C(g) = 1$, if $u \neq 0$ and $C(g) = 0$, if $u = 0$.*

*Proof.* If $\gcd(d, 2^n - 1) = \gcd(d - 1, 2^n - 1) = 1$ then $\alpha^{d-1}$ is a bijection. Using the lemma we can write the capacity of $g$ as

$$\begin{aligned} C(g) &= \sum_{a \neq 0} \rho(a)^2 = 2^{-2n} \sum_{\alpha \neq 0} \mathcal{W}'_{\mathrm{Tr}(f)}(\alpha^{d-1}u)^2 \\ &= 2^{-2n} \sum_{\alpha \neq 0} \mathcal{W}'_{\mathrm{Tr}(f)}(\alpha)^2 = 2^{-2n} \sum_{\alpha \in V_m} \mathcal{W}'_{\mathrm{Tr}(f)}(\alpha)^2 = 1, \end{aligned}$$

where we used the one-to-one correspondence between $a$ and $\alpha$, the fact that $\mathcal{W}'_{\mathrm{Tr}(f)}(0) = 0$ and Parseval's theorem. □

Let us now consider the known examples of highly nonlinear functions $f(x) = x^{-1}$ and $f(x) = x^3$. The former satisfies the conditions of Theorem 4.6 always, while the latter satisfies them only for odd $n$. We get the following corollary.

**Corollary 4.7.** *Let $f(x) = x^d$ in $\mathbb{F}_{2^n}$, where $d = -1$ and $n$ is arbitrary, or $d = 3$ and $n$ is odd. Then the linear approximation $g(x) = x^d + ux$ has capacity $C(g) = 1$, if $u \neq 0$ and $C(g) = 0$, if $u = 0$.*

Hence, increasing dimension $n$ will make $x^{-1}$ more resistant against one-dimensional linear attacks. However, if one uses $n$ equations the capacity does not depend on $n$ and the resistance against multidimensional linear cryptanalysis is the same for all $n > 1$.

## 5. Capacities of Distributions from Simple Filter Generators

Let us study a simple example of a key stream generator (k.s.g.) consisting of an LFSR with $k$ state blocks of size $l$ bits each, and a filter function $f : V_n \to V_m$, where $n$ is a multiple of $l$ and $m \leq n$. Let the LFSR recursion be $\bigoplus_{j=0}^{k} b_j s_{t+j} = 0$, where $b_j \in V$ and $b_0 = b_k = 1$. At time $t$, let $z_t = f(S_t)$, where the input $S_t$ is some fixed subset of the LFSR state blocks. Then $\bigoplus_{j=0}^{k} b_j S_{t+j} = 0$. Our goal is to determine the distribution of $z = \bigoplus_{j=0}^{k} b_j z_{t+j}$. Let the indices of the non-zero coefficients $b_j$ be $0 = j_1 \leq \cdots \leq j_d = k$, where $d$ is the number of non-zero coefficients. Denote $x_i = S_{t+j_i}$. Then $\bigoplus_{i=1}^{d} x_i = 0$. Assume $x_1, \ldots, x_{d-1}$ are statistically independent and uniformly distributed. Given a one-dimensional mask $w \in V_m$, then the correlation $c(w)$ between $w \cdot z$ and 0 can be calculated by

$$\begin{aligned}
c(w) &= c(w \cdot \bigoplus_{j=0}^{k} b_j z_{t+j}) = c\left( w \cdot (\bigoplus_{i=1}^{d-1} f(x_i) + f(\bigoplus_{i=1}^{d-1} x_i)) \right) \\
&= 2^{-(d-1)n} \sum_{x_1,\ldots,x_{d-1} \in V_n} (-1)^{w \cdot (f(x_1)+\cdots+f(x_{d-1})+f(x_1+\cdots+x_{d-1}))} \qquad (10) \\
&= 2^{-dn} \sum_{x_1,\ldots,x_d \in V_n} (-1)^{w \cdot (f(x_1)+\cdots+f(x_d))} \sum_{u \in V_n} (-1)^{u \cdot (x_1+\cdots+x_d)} \\
&= 2^{-dn} \sum_{u \in V_n} \widehat{w \cdot f}(u)^d. \qquad (11)
\end{aligned}$$

Formula (11) was proved in [6]. It can also be considered as a variation of general correlation theorems given in [11]. The one-dimensional masks $w_i$ will be used to construct an $m \times m$ mask matrix $W$. Using the Boolean function

$$(x_1, \ldots, x_{d-1}) \to W \cdot \left( \bigoplus_{i=1}^{d-1} f(x_i) + f(\bigoplus_{i=1}^{d-1} x_i) \right) \tag{12}$$

one can launch a distinguishing attack, where the data complexity is inversely proportional to the capacity of this key stream approximation.

We are going to study three examples: $f(x) = x^{-1}$, $f(x) = x^3$ and $f$ is any bent function. Formula (11) is not useful for functions with complex Walsh spectrum. Therefore, we will use (10) directly, and restrict to the special case where $d = 3$ for the first two examples. Then the recursion equation becomes $s_t + s_{t+i} + s_k = 0$, for some $0 < i < k$. We will denote $x_1$ and $x_2$ by $x$ and $y$, respectively. Then (10) becomes

$$c(w) = 2^{-2n} \sum_{x,y \in V_n} (-1)^{w \cdot (f(x)+f(y)+f(x+y))}. \tag{13}$$

### 5.1. **Filter Function based on the S-box of AES**

It is sufficient to study $x^{-1}$ instead of the actual S-box as making a linear transformation of full rank of a function does not affect its statistical properties.

**Theorem 5.1.** *Let $f : V_n \to V_m$ be the filter function of the k.s.g. described above obtained from the function $x^{-1}$ in $F_{2^n}$ by truncating its output to m bits. Then the correlations $c(w)$ are the same for all $w \neq 0$. Moreover, for any invertible $m \times m$ output mask $W$, the capacity of distinguisher* (12) *is*

$$C = \begin{cases} (2^m - 1)2^{4-2n}, & \text{if } n \text{ even} \\ (2^m - 1)2^{2-2n}, & \text{if } n \text{ odd}. \end{cases}$$

*Proof.* We extend $w \in V_m$ to $V_n$ by appending zeros to it if necessary. We can write (13) as $2^{2n}c(w) = \sum_\eta (-1)^{w \cdot \eta} N_\eta$, where $N_\eta = \#\{(x, y) | x^{-1} + y^{-1} + (x + y)^{-1} = \eta\}$. We divide the sum over $\eta$ to two parts depending on whether $\eta = 0$ or $\eta \neq 0$. Let us calculate $N_0$ first. If $x = 0$ or $y = 0$ or $x + y = 0$ then $\eta = 0$. Hence, assume that $x \neq 0$, $y \neq 0$ and

$x + y \neq 0$. Then the following equivalences hold:

$$\eta = 0 \Leftrightarrow \frac{1}{x} + \frac{1}{y} + \frac{1}{x+y} = 0 \Leftrightarrow \left(\frac{x}{y}\right)^2 + \frac{x}{y} + 1 = 0.$$

The last equation has either 0 or 2 solutions $x$, for every $y \neq 0$, and it has two solutions if and only if $\mathrm{Tr}(1) = 0$ that is, when $n$ is even. Hence,

$$\hat{N}_0 = \#\{(x,y) | x \neq 0,\ y \neq 0,\ x + y \neq 0, \eta = 0\} = \begin{cases} 2(2^n - 1), & \text{if } n \text{ even} \\ 0, & \text{otherwise.} \end{cases}$$

Then $N_0$ is given by

$$\begin{aligned} N_0 &= \#\{(x,y) | x = 0, \eta = 0\} + \#\{(x,y) | x \neq 0, y = 0, \eta = 0\} \\ &\quad + \#\{(x,y) | x, y \neq 0, x = y, \eta = 0\} + \hat{N}_0 \\ &= \begin{cases} 5 \cdot 2^n - 4, & \text{if } n \text{ even} \\ 3 \cdot 2^n - 2, & \text{otherwise.} \end{cases} \end{aligned}$$

Now we calculate the $N_\eta$ for $\eta \neq 0$. Then $x \neq 0$, $y \neq 0$ and $x \neq y$, and consequently, $y \neq \frac{1}{\eta}$. We have that $f(x) + f(y) + f(x+y) = \eta$ if and only if $\left(\frac{x}{y}\right)^2 + \frac{x}{y} + \frac{1}{1+\eta y} = 0$. This has two solutions $x$ for each $y \neq 0$ if and only if $\mathrm{Tr}(\frac{1}{1+\eta y}) = 0$. Note that the solutions must also satisfy $x \neq 0$ and $x \neq y$. The function $y \mapsto 1/(1 + \eta y)$ defined from $\mathbb{F}_{2^n} \setminus \{\eta^{-1}, 0\} \to \mathbb{F}_{2^n} \setminus \{0, 1\}$ is a bijection. Trace is a linear mapping so $\#\{y | \mathrm{Tr}(y) = 0\} = 2^{n-1}$. We obtain $\frac{1}{2}N_\eta = \#\{y \neq 0, y \neq \eta^{-1} \mid \mathrm{Tr}(1/(1 + \eta y)) = 0\}$.

If $n$ is odd, then $N_\eta = 2^n - 2$, since the function $\mathrm{Tr}(1/(1+\eta y)) \neq 0$ for $y = \eta^{-1}$. If $n$ is even, then both values which this function never takes, have trace equal to zero, and we obtain $N_\eta = 2^n - 4$.

Combining the results we get that the correlation for all non-zero one-dimensional output masks $w$ is $c(w) = 2^{2-n}$, if $n$ is even and $c(w) = 2^{1-n}$, if $n$ is odd. The claim follows now from Corollary 3.2. □

By choosing any $n$ linearly independent one-dimensional masks, for example, $e_i$, $i = 1, \ldots, n$, as the base masks, we can make an optimal multidimensional distinguisher. This example also shows that the assumption about statistically independent base approximations cannot hold even though the selected masks $(e_i)$ are linearly independent. This follows since the combination correlations are of the same magnitude as the base correlations.

### 5.2. Function $x^3$ as a Filter Function

**Theorem 5.2.** *Let $f : V_n \to V_m$ be the filter function of the k.s.g. described above obtained from function $x^3$ in $F_{2^n}$ by truncating its output to m bits. Then the correlations $c(w)$ are the same for all $w \neq 0$ and the capacity of the distinguisher* (12) *is $C = (2^m - 1)2^{2-2n}$ for any invertible mask $W$.*

*Proof.* We extend $w \in V_m$ to $V_n$ by appending zeroes to it if necessary. Using (13) we get

$$c(w) = 2^{-2n} \sum_{x,y} (-1)^{w\cdot(x^2y+y^2x)} = 2^{-n} + 2^{-2n} \sum_{y\neq 0} \sum_{x} (-1)^{w\cdot(x^2y+y^2x)}.$$

For all $y \neq 0$, $x^2y + y^2x$ is linear. It goes twice through the values of an $n-1$-dimensional vector subspace $V_{n-1}(y) := \{x^2y + y^2x | x \in V_n\}$. Since $V_{n-1}(y)$ is a vector subspace, there exists $c \neq 0$, such that $c \perp z$ for all $z \in V_{n-1}(y)$. Moreover, if $y_1 \neq y_2$, then $c_1 \neq c_2$. Then for any $w \neq 0$ there exists a unique $y_w$ such that $w \cdot (x^2y_w + y_w^2x) = 0$. If $y \neq y_w$, $w \cdot (x^2y + y^2x)$ takes the values 0 and 1 equally many times. Hence,

$$c(w) = 2^{-n} + 2^{-2n} \left( \sum_{y=y_w} \sum_{x} (-1)^{w\cdot(x^2y+y^2x)} + \sum_{y\neq y_w,0} \sum_{x} (-1)^{w\cdot(x^2y+y^2x)} \right) = 2^{1-n}.$$

By Corollary 3.2, we get the claim. □

### 5.3. A Bent Function as a Filter Function

In this case we can study the general case of an LFSR with $d \geq 3$. Suppose that $f : V_n \to V_m$ is bent.

**Theorem 5.3.** *Let $f(x)$ be a bent filter function of the k.s.g. described above. Then, for any fixed even $d \geq 4$, the correlations $c(w)$ are the same, for all $w \neq 0$, and for any fixed odd $d \geq 3$ the absolute values of the correlations $|c(w)|$ are the same, for all $w \neq 0$. The capacity of the distinguisher* (12) *is $C = \sum_{a\neq 0} 2^{-2n} = (2^m - 1)2^{-2n}$ for any invertible mask $W$.*

*Proof.* Denote the right hand side of (10) by $c_d$. If $d = 3$, then using (13) we get

$$c_3 = 2^{-2n} \sum_{x} (-1)^{w\cdot f(x)} \sum_{y} (-1)^{w\cdot(f(y)+f(x+y))}.$$

Since $f$ is bent we know that the sum over $y$ is $2^n$ if $x = 0$ and zero otherwise. Hence, $c_3 = 2^{-n}(-1)^{w\cdot f(0)}$. If $d = 4$, we obtain similarly that $c_4 = 2^{-n}\sum_x(-1)^{w\cdot(f(x)+f(x))} = 2^{-n}$. Finally, for $d \geq 5$, we get

$$\begin{aligned} c_d &= 2^{-(d-1)n} \sum_{x_1,\ldots,x_{d-2}} (-1)^{w\cdot(f(x_1)+\cdots+f(x_{d-2}))} \sum_{x_{d-1}} (-1)^{w\cdot(f(x_{d-1})+f(x_1+\cdots+x_{d-1}))} \\ &= 2^{-(d-2)n} \sum_{x_1,\ldots,x_{d-3}} (-1)^{w\cdot(f(x_1)+\cdots+f(x_{d-3})+f(x_1+\cdots+x_{d-3}))} = c_{d-2}. \end{aligned}$$

Therefore, $c_d = (-1)^{w\cdot f(0)}2^{-n}$, if $d$ is odd and $2^{-n}$ if $d$ is even. Since $c(w)^2 = c_d^2 = 2^{-2n}$ for all $w \neq 0$, we get the claim. □

In all three examples, the amount of data needed for the attack is approximately $2^{2n-m}$. If $n = 2m$, a multidimensional distinguishing attack takes $2^{3m}$ words of data whereas the one-dimensional attack needs $2^{4m}$ words of data.

## 6. Constructing Multidimensional Linear Approximations

One approach to constructing multidimensional linear approximations is to search for strong one-dimensional linear approximations over the whole system and use Lemma 2.1 to construct the multidimensional approximations. Since many linear distinguishing attacks have already concentrated on finding the best one-dimensional mask, this can be seen as a practical way to apply the theory, in particular, when several approximations about equally large correlation have been found, see e.g. [8]. A second approach, taken in [4], is to consider p.d. of linear approximation of each part of the system separately and then use convolution to obtain the p.d. of the entire linear approximation.

We say that two Boolean functions are statistically independent if they do not share common inputs. The generalisation of the Piling-up lemma can be stated as follows:

**Lemma 6.1.** *Suppose that $g$ and $h$ are statistically independent. Then $p(g + h) = p(g) * p(h)$.*

In the one-dimensional case, the capacity of the convolution is the product of the original capacities. Unfortunately, this useful fact cannot be generalised to multiple dimensions, where only the following result holds in general.

**Theorem 6.2.** *Let g and h be statistically independent. Then*

$$C(g+h) \leq C(g)C(h). \tag{14}$$

*Proof.* Denote the correlation $\rho(a) = c(a \cdot g)$ corresponding to function $g$ by $\rho_g(a)$. Using (3) and (1) we get $\rho_{g+h}(a) = \rho_g(a)\rho_h(a)$, for all $a \neq 0$. To compute the capacity of $g + h$ we use Corollary 3.2 to get

$$C(g+h) = \sum_{a\neq 0} \rho_{g+h}(a)^2 \leq \sum_{a\neq 0} \rho_g(a)^2 \sum_{a\neq 0} \rho_h(a)^2 = C(g)C(h).$$

□

If the inputs of the two functions to be combined are statistically independent, we can use the *m*-dimensional Piling-up lemma 6.1 to calculate the p.d. over the whole system. The inputs are often assumed to be statistically independent, for example, when combining rounds of a block cipher. However, due to the linear hull effect [12], this is just an approximation. In the following theorem, we give the distribution of a linear approximation over a composition of two Boolean functions $f_1$ and $f_2$ in terms of the distributions of the linear approximations of $f_1$ and $f_2$, thus generalising the basic correlation formula of [13].

**Theorem 6.3** (Correlation theorem)**.** *Let $f_1 : V_l \to V_n$, $f_2 : V_n \to V_k$ and let $g_V = Vf_1 + U$ and $h_V = Wf_2 + V$, where $U \in V_l^m$, $V \in V_n^m$ and $W \in V_k^m$. Let $f = W(f_2 \circ f_1) + U$. Then for all matrices U and W*

$$p(f) = 2^{-mn+n} \sum_{V\in V_n^m} p(g_V) * p(h_V) - (2^n - 1)\theta_m. \tag{15}$$

*Proof.* Let us start by studying the sum over $V$ using the multi-Walsh transform as follows:

$$\sum_{V\in V_n^m} \mathcal{W}_{g_V}(x)\mathcal{W}_{h_V}(x) = 2^{-n-l} \sum_{\xi\in V_l} \sum_{\zeta\in V_n} \prod_i x_i^{u_i\cdot\xi\oplus w_i\cdot f_2(\zeta)} \sum_{V\in V_n^m} \prod_i x_i^{v_i\cdot(f_1(\xi)\oplus\zeta)}.$$

The sum is now divided to two parts, $S_1$ and $S_2$, where $S_1$ is the sum with $\zeta = f_1(\xi)$ and $S_2$ is the sum where $\zeta \neq f_1(\xi)$. If $a = 0$, then $\sum_{V\in V_n^m} \prod_i x_i^{v_i\cdot a} = 2^{nm}$. On the other hand, if $a \neq 0$, then the sum is $2^{nm}\Theta_m(x)$. Using this we get

$$S_1 = 2^{-n-l} \sum_{\xi} \prod_i x_i^{u_i\cdot\xi\oplus w_i\cdot f_2(f_1(\xi))} 2^{nm} = 2^{-n+mn}\mathcal{W}_f(x)$$

$$S_2 = 2^{-n-l} \sum_{\xi} \sum_{\zeta\neq f_1(\xi)} \prod_i x_i^{u_i\cdot\xi\oplus w_i\cdot f_2(\zeta)} 2^{nm}\Theta_m(x) = 2^{nm-n}\,(2^n - 1)\,\Theta_m(x).$$

Solving $\mathcal{W}_f(x)$ from these equations and using (7) gives the desired result. □

Similarly to the proof of Theorem 6.2, by using (15), we can prove that

$$C(f) \leq 2^{-2mn+2n+m} \sum_{V \in V_n^m} C(g_V)C(h_V). \tag{16}$$

In Sections 4 and 5 we saw that functions that are considered strong against one-dimensional cryptanalysis can actually be quite weak against multidimensional cryptanalysis. On the other hand, we do not know how much the capacities are weakened when such strong multidimensional approximations are combined over multiple rounds as (14) and (16) give just upper bounds.

## 7. Conclusions

We investigated theoretical aspects of multidimensional linear distinguishing attacks on concrete examples of Boolean functions. First, we showed how the probability distribution of a multidimensional linear approximation can be determined based on the correlations of the one-dimensional linear approximations. Secondly, we saw that significant reduction in data complexity compared to one-dimensional linear distinguishers can be achieved. Thirdly, we presented a number of concrete, albeit theoretical examples where one cannot find even a single pair of statistically independent one-dimensional linear approximations, which shows that the presumptions adopted in [2], [3] and [7] are not valid in general.

## Acknowledgement

We wish to thank Gregor Leander for noticing a problem in our results and proofs in Section 4.2 and for proposing a new elegant solution.

## References

[1] Matsui, M.: Linear cryptanalysis method for DES cipher. In: EUROCRYPT '93: Workshop on the theory and application of cryptographic techniques on Advances in cryptology, Secaucus, NJ, USA, Springer-VerlagNew York, Inc. (1993) 386–397

[2] Burton S. Kaliski, J., Robshaw, M.J.B.: Linear Cryptanalysis Using Multiple Approximations. In: CRYPTO '94: Proceedings of the 14th Annual International Cryptology Conference on Advances in Cryptology, London, UK, Springer-Verlag (1994) 26–39
[3] Biryukov, A., Cannire, C.D., Quisquater, M.: On Multiple Linear Approximations. In: Advances in Cryptology - CRYPTO 2004. Volume 3152 of Lecture Notes in Computer Science., Springer-Verlag (2004) 1–22
[4] Englund, H., Maximov, A.: Attack the Dragon. In: Progress in Cryptology - INDOCRYPT 2005. Volume 3797 of Lecture Notes in Computer Science. (2005) 130–142
[5] Baignères, T., Junod, P., Vaudenay, S.: How Far Can We Go Beyond Linear Cryptanalysis? In: Advances in Cryptology - ASIACRYPT 2004, 10th International Conference on the Theory and Application of Cryptology and Information Security. Volume 3329 of Lecture Notes in Computer Science. (2004) 432–450
[6] Molland, H., Helleseth, T.: An improved correlation attack against irregular clocked and filtered keystream generators. In: Advances in Cryptology-CRYPTO 2004. Volume 3152., Springer-Verlag (2004) 373–389
[7] Murphy, S.: The Independence of Linear Approximations in Symmetric Cryptology. IEEE Transactions on Information Theory **27** (2006) 5510–5518
[8] Hermelin, M., Nyberg, K., Cho, J.Y.: Multidimensional Linear Cryptanalysis of Reduced Round Serpent. In Yi Mu, Willy Susilo, J.S., ed.: Information Security and Privacy: The 13th Australasian Conference on Information Security and Privacy. Lecture Notes in Computer Science, Springer-Verlag (2008)
[9] Cover, T.M., Thomas, J.A.: Elements of Information Theory. 2nd edn. Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience (2006)
[10] Nyberg, K., Hermelin, M.: Multidimensional Walsh Transform and a Characterization of Bent Functions. In Tor Helleseth, P.V.K., Ytrehus, O., eds.: Proceedings of the 2007 IEEE Information Theory Workshop on Information Theory for Wireless Networks. IEEE (2007) 83–86
[11] Nyberg, K.: Correlation theorems in cryptanalysis. Discrete Applied Mathematics (2000)
[12] Nyberg, K.: Linear approximation of block ciphers. In: Advances in Cryptology - EUROCRYPT'94. Volume 950. (1995)
[13] Daemen, J., Govaerts, R., Vandewalle, J.: Correlation matrices. In Preneel, B., ed.: Fast Software Encryption. Volume 1008 of Lecture Notes in Computer Science., Springer-Verlag (1995) 275–285