Counting and characterising functions with “fast points” for differential attacks

Cryptogr. Commun. (2017) 9:217–239
DOI 10.1007/s12095-015-0166-1

Counting and characterising functions with "fast points" for differential attacks

Ana Sălăgean[1] · **Matei Mandache-Sălăgean**[2]

Received: 22 April 2015 / Accepted: 21 October 2015 / Published online: 26 November 2015

Abstract Higher order derivatives have been introduced by Lai in a cryptographic context. A number of attacks such as differential cryptanalysis, the cube and the AIDA attack have been reformulated using higher order derivatives. Duan and Lai have introduced the notion of "fast points" of a polynomial function f as being vectors $\mathbf{a}$ so that computing the derivative with respect to $\mathbf{a}$ decreases the total degree of f by more than one. This notion is motivated by the fact that most of the attacks become more efficient if they use fast points. Duan and Lai gave a characterisation of fast points and Duan et al. gave some results regarding the number of functions with fast points in some particular cases. We firstly give an alternative characterisation of fast points and secondly give an explicit formula for the number of functions with fast points for any given degree and number of variables, thus covering all the cases left open in Duan et al. Our main tool is an invertible linear change of coordinates which transforms the higher order derivative with respect to an arbitrary set of linearly independent vectors into the higher order derivative with respect to a set of vectors in the canonical basis. Finally we discuss the cryptographic significance of our results.

Keywords Higher order differential attacks · Higher order derivative · Cryptanalysis · Polynomials over finite fields

Mathematics Subject Classification (2010) 94A60 · 11T55

✉ Ana Sălăgean
a.m.salagean@lboro.ac.uk

Matei Mandache-Sălăgean
mfm41@cam.ac.uk

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

[2] Trinity College, University of Cambridge, Cambridge, UK

1 Introduction

Higher order derivatives were introduced in a cryptographic context by Lai in [6]. This notion had already been used for a very long time, under the name of finite difference (see also discrete derivative, Delta operator, difference equations etc.), in other areas of mathematics (notably for the numerical approximation of the derivative, or as a discrete analogue of the derivative). The discrete derivative of a function f with respect to a vector $\mathbf{a}$ is the function $\Delta_{\mathbf{a}} f(\mathbf{x}) = f(\mathbf{x}+\mathbf{a}) - f(\mathbf{x})$. A higher order derivative of order k is obtained by repeated application of this operator, k times, with respect to k vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$.

A number of cryptographic attacks can be reformulated using higher order derivatives. Differential cryptanalysis (introduced by Biham and Shamir [1]) has been thus reformulated by Lai in [6]; the cube attack of Dinur and Shamir [2] and the related AIDA attack of Vielhaber [7] have been reformulated in Knellwolf and Meier [5], Duan and Lai [3]. Other attacks are also mentioned in [3].

Most attacks mentioned above treat the cryptographic function as a "black box" boolean function f. Computing a higher order derivative of order k will involve 2^k calls to the "black box" function f. Any boolean function can be represented in Algebraic Normal Form, i.e. as a polynomial over $\mathbb{F}_2$ with degree at most one in each variable. Differentiation decreases the total degree d of a function by at least one. The attacks rely on computing higher order derivatives to obtain a function which has some "non-random" behaviour, ideally it is a linear function. If the degree decreases by exactly one for each differentiation, then we would need differentiation of order $k = d - 1$ (i.e. 2^{d-1} calls to f) to achieve a linear function. Most well designed cryptographic functions have a high degree, so 2^{d-1} would be prohibitively large. However, if the degree decreases by more than one for some of the differentiation steps, an attack can be successful for an order of differentiation k considerably lower then $d - 1$.

Motivated by this, Duan and Lai [3] introduced the notion of "fast points". Namely $\mathbf{a}$ is a fast point for a polynomial function f if differentiation with respect to $\mathbf{a}$ decreases the degree of f by two or more.

In [3], Duan and Lai gave characterisations of fast points. In [4], Duan et al. showed that given a function f, its set of fast points forms a vector space. They also started to investigate the number of polynomial functions with fast points among the polynomial function of given degree d in n variables. They succeeded giving exact formulae for a few particular cases (degree 1,2, $n-2$, $n-1$); for the other degrees, numerical results were given when the number of variables is small (at most 8) by exhaustively enumerating all these functions and checking whether they have fast points. Such an approach is very computationally intensive (more than exponential in the degree of the polynomial) so some values are missing in their tables for 7 and 8 variables.

We continue the work commenced in [3] and [4] using a different approach. Our main tool is a suitably chosen linear change of variables. While we are mostly interested in results over the binary field, when possible we will formulate results more generally, for finite fields or for arbitrary fields. We show that differentiation with respect to an arbitrary set of vectors can be transformed, via an invertible linear change of variables, into differentiation with respect to a set of vectors in the canonical basis. (For an analogy, in the case of functions in several variables over the real numbers, directional derivatives can be transformed into partial derivatives via a suitable change of coordinates.) It is much easier to characterise and

to count functions which admit canonical basis vectors as fast points. We then transfer these results to functions that have arbitrary fast points.

We give thus an alternative characterisation of fast points (Corollary 2). Let f be a function in n variables and of degree d. In essence, we show that f has fast points if, when ignoring its monomials of degree less than d, f is actually a function in less than n variables, possibly "disguised" by an invertible linear change of coordinates. For example, the function $f(x_1, x_2, x_3) = x_1x_2 + x_1x_3$ looks like a function in 3 variables, but can actually be viewed as a function in two variables, $g(y_1, y_2) = y_1y_2$ followed by the change of variables $y_1 = x_1$, $y_2 = x_2 + x_3$. When designing cryptographic functions, one should therefore avoid such functions, as much as possible.

Next, in Section 5 we count the number $|F(n, d)|$ of functions of degree d in n variables that have fast points. We obtain a recurrence relation for $|F(n, d)|$, and also an explicit formula (Theorem 6). We further refine our results to give within each degree d and number of variables n, the number of functions whose fast points form a space of dimension k. Numerical values can then easily be computed at minimal computational cost (the number of integer multiplications/additions is polynomial in n). For illustration, in Section 7 we displayed the results for up to 8 variables, thus filling in the gaps in the table of Duan et al. [4].

The effect of a change of variables on higher order differentiation is examined in Section 6. We propose a natural generalisation of "fast points" to "fast spaces". Counting functions with fast spaces is probably feasible but rather difficult and of less interest, so we have not pursued it further.

Finally we discuss in Section 8 the cryptographic significance of our results. Using our previous results we estimate probabilities of a function having fast points and also give some asymptotic results. Perhaps not surprisingly, it turns out that fast points are relatively rare. For $3 \leq d \leq n-3$, the proportion of functions of degree d with fast points out of the total number of functions of degree d decreases very fast with n and is asymptotically zero. The ratio is approximately $\frac{1}{2^{\binom{n-1}{d-1}-n}}$, so its decrease, as n increases, is faster than conjectured in [4, Remark 12] (their conjecture was a geometric series in n). For a given number of variables n, polynomials of degree approximately $n/2$ are least likely to have fast points; this probability increases as we move towards lower or higher degrees. We also show that for a fixed n and d (again $3 \leq d \leq n-3$) out of the functions that do have fast points, most have only one fast point, then much fewer have 3 fast points, even fewer have $2^3 - 1$ fast points etc. One should keep in mind though that these results assume that the function is picked uniformly at random. In practice, cryptographic functions have additional constraints, and an attacker should exploit such information to improve their chances of finding fast points.

2 Preliminaries

Throughout this paper K will denote an arbitrary field and $\mathbb{F}_p$ denotes the finite field with p elements where p is a prime. We denote by $\mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0) \in K^n$ the vector which has a 1 in position i and zeroes elsewhere, i.e. $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is the canonical basis of the vector space K^n. We will denote by $\langle \mathbf{a}_1, \ldots, \mathbf{a}_n \rangle$ the vector subspace of K^n generated by $\mathbf{a}_1, \ldots, \mathbf{a}_n \in K^n$. The all-zero vector will be denoted by $\mathbf{0}$.

We recall the definition of (discrete) derivative/ differentiation here:

Definition 1 Let $f : K^n \to K$ be a function in n variables $x_1, \ldots, x_n$. Let $\mathbf{a} = (a_1, \ldots, a_n) \in K^n \setminus \{\mathbf{0}\}$. The *differentiation* operator with respect to a vector $\mathbf{a}$ associates to each function f its *discrete derivative* $\Delta_{\mathbf{a}} f$ defined as

$$\Delta_{\mathbf{a}} f(x_1, \ldots, x_n) = f(x_1 + a_1, \ldots, x_n + a_n) - f(x_1, \ldots, x_n).$$

Denoting $\mathbf{x} = (x_1, \ldots, x_n)$ we can also write $\Delta_{\mathbf{a}} f(\mathbf{x}) = f(\mathbf{x} + \mathbf{a}) - f(\mathbf{x})$.

For the particular case of $a = \mathbf{e}_i$ for some $1 \leq i \leq n$, we will call $\Delta_{\mathbf{e}_i}$ differentiation (or discrete derivative) w.r.t. the variable x_i.

Remark 1 Note that the discrete derivative with respect to a variable x should not be confused with the formal derivative with respect to x. The two notions coincide for polynomials of degree at most one in x, but they are different for higher degrees. For example, for the function $f : \mathbb{F}_5 \to \mathbb{F}_5$, $f(x) = x^3$, the discrete derivative is $3x^2 + 3x + 1$ whereas the formal derivative is $3x^2$. Functions over $\mathbb{F}_2$ can always be represented as polynomial functions of degree at most one in each variable, so in this case the two notions coincide.

The differentiation operator is a linear operator. Repeated application of this operator (which is commutative and associative) is called higher order differentiation and the result of applying it to a function will be called higher order derivative. It will be denoted by

$$\Delta^{(k)}_{\mathbf{a}_1, \ldots, \mathbf{a}_k} f = \Delta_{\mathbf{a}_1} \Delta_{\mathbf{a}_2} \ldots \Delta_{\mathbf{a}_k} f$$

where $\mathbf{a}_1, \ldots, \mathbf{a}_k \in K^n$ are not necessarily distinct.

The change of variables is a widely used technique in different areas of mathematics. Here we are only interested in invertible linear changes of variables (which can also be viewed as a linear change of coordinates, or a change of basis).

Definition 2 Let T be an invertible $n \times n$ matrix over K. We say that T defines the *invertible linear change of variables* described as $\mathbf{x} = T\mathbf{y}$.

If $f : K^n \to K$ is a function in n variables $x_1, \ldots, x_n$, the function obtained from f by this change of variables is the function $g(\mathbf{y}) = f(T\mathbf{y})$. In other words, defining $\varphi_T : K^n \to K^n$ as $\varphi(\mathbf{y}) = T\mathbf{y}$, the change of variable defined by T is the composition by φ_T, i.e. $f \circ \varphi_T$.

Note that in the definition above the change of variables is indeed invertible: the function f can also be obtained from the function g by the linear change of variable $f(\mathbf{x}) = g(T^{-1}\mathbf{x})$. In other words, $(\varphi_T)^{-1} = \varphi_{T^{-1}}$.

If f is a polynomial, we will denote by $\deg(f)$ the total degree of f, with the usual convention of $\deg(0) = -\infty$. The following is a well known result needed later:

Proposition 1 *The total degree of a polynomial is preserved under invertible linear changes of variables.*

(For a quick proof, note first that the degree cannot increase after a linear change of coordinate. If $g(\mathbf{y}) = f(T\mathbf{y})$ then $\deg(g) \leq \deg(f)$. On the other hand, we can view f as being obtained from g via the linear change of variables given by T^{-1}, so by the same argument $\deg(f) \leq \deg(g)$.)

Differentiation decreases the degree of a polynomial by at least one:

Proposition 2 [6] *Let* $f : K^n \to K$ *be a polynomial function in* n *variables and* $\mathbf{a} \in K^n \setminus \{\mathbf{0}\}$. *Then* $\deg(\Delta_{\mathbf{a}} f) \leq \deg(f) - 1$.

We will be interested in the situations where the degree decreases by more than one.

Definition 3 [3, 4] Let $f : K^n \to K$ be a non-constant polynomial function in n variables and $\mathbf{a} \in K^n \setminus \{\mathbf{0}\}$. We call $\mathbf{a}$ a *fast point* for f if $\deg(\Delta_{\mathbf{a}} f) < \deg(f) - 1$.

For convenience we will also consider $\mathbf{0}$ as a (trivial) fast point for any polynomial function f (since $\Delta_{\mathbf{0}} f = f(\mathbf{x} + \mathbf{0}) - f(\mathbf{x}) = 0$, although we do not normally define the differentiation operator for the difference $\mathbf{0}$).

It was shown in [4] that over $\mathbb{F}_2$ the set of fast points for a polynomial function f is a linear space. The result actually holds over arbitrary fields:

Theorem 1 *(cf.* [4, Lemma 3.1] *) Let* $f : K^n \to K$ *be a polynomial function. The set of all fast points of* f *is a linear space.*

The proof is the same as in [4], namely putting $\mathbf{a} = (a_1, \ldots, a_n)$ and $\deg(f) = d$, the coefficient of each of the monomials of degree $d - 1$ in $\Delta_{\mathbf{a}} f$ is a linear expression in $a_1, \ldots, a_n$. The vector $\mathbf{a}$ is a fast point iff all those expressions equal zero, i.e. $a_1, \ldots, a_n$ is a solution of the corresponding linear system of equations. In the binary case, there are $\binom{n}{d-1}$ terms of degree $d - 1$, so in general there are $\binom{n}{d-1}$ equations (but depending on f, some of these equations can be degenerate $0 = 0$).

Recall that all functions $f : \mathbb{F}_p^n \to \mathbb{F}_p$ can be uniquely represented in Algebraic Normal Form, i.e. as a polynomial function corresponding to a polynomial of degree at most $p - 1$ in each variable.

We will need to count the number of vector subspaces of a given vector space. We recall the notion of Gaussian binomial coefficients (the definition can be given in a more general form, but we only need the form below):

Definition 4 Let $0 \leq k \leq n$ and $q > 1$ be integers. The Gaussian binomial coefficients (or q-binomial coefficients) are defined as

$$\binom{n}{k}_q = \frac{(q^n - 1)(q^{n-1} - 1) \cdots (q^{n-k+1} - 1)}{(q^k - 1)(q^{k-1} - 1) \cdots (q - 1)}.$$

Proposition 3 *The number of vector subspaces of dimension* k *of the vector space* $\mathbb{F}_q^n$ *is equal to* $\binom{n}{k}_q$.

3 Change of variables and differentiation

We first show that differentiation with respect to a vector $\mathbf{a}$ can be transformed, using a suitable change of variables, to differentiation w.r.t. a canonical basis vector $\mathbf{e}_j$, i.e. differentiation w.r.t. one variable x_j.

Theorem 2 *Let f be a function of n variables $f : K^n \to K$ and let T be an invertible $n \times n$ matrix over K. Denote by g the function obtained from f via the change of variables defined by T, namely: $g(\mathbf{y}) = f(T\mathbf{y})$. Then for any $\mathbf{a} \in K^n \setminus \{\mathbf{0}\}$ we have:*

$$(\Delta_{\mathbf{a}} f)(T\mathbf{y}) = (\Delta_{T^{-1}\mathbf{a}} g)(\mathbf{y}).$$

In particular, if $\mathbf{a}$ equals column j of T, we have:

$$(\Delta_{\mathbf{a}} f)(T\mathbf{y}) = (\Delta_{\mathbf{e}_j} g)(\mathbf{y})$$

or equivalently

$$(\Delta_{\mathbf{a}} f)(\mathbf{x}) = (\Delta_{\mathbf{e}_j} g)(T^{-1}\mathbf{x}).$$

Proof Applying the change of variables T to $(\Delta_{\mathbf{a}} f)(\mathbf{x}) = f(\mathbf{x} + \mathbf{a}) - f(\mathbf{x})$ we obtain:

$$\begin{aligned}(\Delta_{\mathbf{a}} f)(T\mathbf{y}) &= f(T\mathbf{y} + \mathbf{a}) - f(T\mathbf{y}) \\ &= f(T(\mathbf{y} + T^{-1}\mathbf{a})) - f(T\mathbf{y}) \\ &= g(\mathbf{y} + T^{-1}\mathbf{a}) - g(\mathbf{y}) \\ &= (\Delta_{T^{-1}\mathbf{a}} g)(\mathbf{y}).\end{aligned}$$

For the case that $\mathbf{a}$ equals column j of T, we have $T\mathbf{e}_j = \mathbf{a}$, so $T^{-1}\mathbf{a} = \mathbf{e}_j$. □

An equivalent formulation of the Theorem above is to say that the following diagram commutes:

$$\begin{array}{ccc} \mathcal{F}(K^n, K) & \xrightarrow{\Delta_{\mathbf{a}}} & \mathcal{F}(K^n, K) \\ \Phi_T \downarrow & & \downarrow \Phi_T \\ \mathcal{F}(K^n, K) & \xrightarrow{\Delta_{\mathbf{e}_j}} & \mathcal{F}(K^n, K) \end{array}$$

where $\mathcal{F}(K^n, K)$ denotes the set of functions from K^n to K and Φ_T denotes the operator of change of variables defined by T, i.e. $\Phi_T(f) = f \circ \varphi_T$.

An important particular case of the theorem above is:

Corollary 1 *Let $\mathbf{a} = (a_1, \ldots, a_n) \in K^n \setminus \{\mathbf{0}\}$ and let j be such that $a_j \neq 0$. For any function of n variables $f : K^n \to K$ we have*

$$(\Delta_{\mathbf{a}} f)(x_1, \ldots, x_n) = (\Delta_{\mathbf{e}_j} g)\left(x_1 - \frac{a_1}{a_j}x_j, \ldots, x_{j-1} - \frac{a_{j-1}}{a_j}x_j, \frac{1}{a_j}x_j, x_{j+1} - \frac{a_{j+1}}{a_j}x_j, \ldots, x_n - \frac{a_n}{a_j}x_j\right)$$

where g is the function obtained from f by the change of variables $g(y_1, \ldots, y_n) = f(y_1 + a_1 y_j, \ldots, y_{j-1} + a_{j-1} y_j, a_j y_j, y_{j+1} + a_{j+1} y_j, \ldots, y_n + a_n y_j)$.

In particular, if the field is $K = \mathbb{F}_2$, we have

$$(\Delta_{\mathbf{a}} f)(x_1, \ldots, x_n) = (\Delta_{\mathbf{e}_j} g)(x_1 + a_1 x_j, \ldots, x_{j-1} - a_{j-1} x_j, x_j, x_{j+1} + a_{j+1} x_j, \ldots, x_n + a_n x_j)$$

where g is the function obtained from f by the change of variables $g(y_1, \ldots, y_n) = f(y_1 + a_1 y_j, \ldots, y_{j-1} + a_{j-1} y_j, y_j, y_{j+1} + a_{j+1} y_j, \ldots, y_n + a_n y_j)$.

Proof We apply Theorem 1 for the matrix T consisting of the identity matrix, except that column j is replaced by column $\mathbf{a}$. Note that when the field is $\mathbb{F}_2$ this matrix equals its inverse. □

4 Characterisation of fast points

For differentiation w.r.t. one variable in $\mathbb{F}_p$ it is easy to characterise fast points:

Proposition 4 *Let f be a function of n variables of degree d, $f : \mathbb{F}_p^n \to \mathbb{F}_p$, and let $1 \leq j \leq n$. We have that $\mathbf{e}_j$ is a fast point for f iff none of the monomials of f of degree d contain x_j.*

The total degree of $\Delta_{\mathbf{e}_j} f$ is precisely one less than the highest total degree among the monomials of f that contain x_j.

Proof Let $f = f_1 + f_2$ with all terms in f_1 having degree d and all terms in f_2 having degree $d-1$ or less.

If x_j does not appear in f_1, then obviously $\Delta_{\mathbf{e}_j} f_1 = 0$, so $\deg(\Delta_{\mathbf{e}_j} f) = \deg(\Delta_{\mathbf{e}_j} f_2) \leq \deg(f_2) - 1 \leq d-2$, i.e. $\mathbf{e}_j$ is a fast point for f.

For the reverse implication, assume $\mathbf{e}_j$ is a fast point for f. Without loss of generality, assume $j = 1$. Assume, for a contradiction, that f_1 does contain x_1. Let $f_1 = x_1^{d_1} g_1 + \ldots + x_1^{d_\ell} g_\ell + f_3$, where $d_1, \ldots, d_\ell$ are distinct integers in $\{1, 2, \ldots, p-1\}$ (since in the Algebraic Normal Form the degree in each variable is at most $p-1$), g_j are polynomials in $x_2, \ldots, x_n$ of total degree $d - d_j$ and f_3 is a polynomial in $x_2, \ldots, x_n$. Then $\Delta_{\mathbf{e}_1} x_1^{d_i} g_i = d_i x_1^{d_i - 1} g_i + h_i$ where h_i has degree $d-2$ or less. So $\Delta_{\mathbf{e}_1} f_1 = d_1 x_1^{d_1 - 1} g_1 + \ldots + d_\ell x_1^{d_\ell - 1} g_\ell + h_1 + \ldots + h_\ell$ where $h_1 + \ldots + h_\ell$ has total degree $d-2$ or less. Since $d_i \leq p-1$ for all i, we have $d_i \bmod p \neq 0$. Also, for all $i \neq j$ none of the monomials in $d_i x_1^{d_i - 1} g_i$ can be a monomial in $d_j x_1^{d_j - 1} g_j$ (their degree in x_1 is different), so they cannot cancel out. This means that $\deg(\Delta_{\mathbf{e}_1} f) = d - 1$, i.e. $\mathbf{e}_1$ is not a fast point for f. Contradiction. □

On the other hand, Theorem 2 allows us to transform fast points using a change of variables:

Theorem 3 *Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ and $\mathbf{a} \in \mathbb{F}_p^n \setminus \{\mathbf{0}\}$. Let T be an invertible matrix and let g be obtained from f via the change of variables defined by T, i.e. $g(\mathbf{y}) = f(T\mathbf{y})$. For any $\mathbf{a} \in \mathbb{F}_p^n \setminus \{\mathbf{0}\}$ we have that $\mathbf{a}$ is a fast point for f iff $T^{-1}\mathbf{a}$ is a fast point for g.*

In particular, if $\mathbf{a}$ equals column j of T, we have that $\mathbf{a}$ is a fast point for f iff $\mathbf{e}_j$ is a fast point for g.

Proof Let $d = \deg(f)$. By Proposition 1, $\deg(g) = d$. Using Theorem 2 we have $(\Delta_{\mathbf{a}} f)(\mathbf{x}) = (\Delta_{T^{-1}\mathbf{a}} g)(T^{-1}\mathbf{x})$. We have that $\mathbf{a}$ is a fast point for f iff $\deg(\Delta_{\mathbf{a}} f) < d - 1$. But the degree in $\mathbf{x}$ of $(\Delta_{\mathbf{a}} f)(\mathbf{x})$ equals the degree in $\mathbf{x}$ of $(\Delta_{T^{-1}\mathbf{a}} g)(T^{-1}\mathbf{x})$. By Proposition 1, the latter equals the degree in $\mathbf{y}$ of $(\Delta_{T^{-1}\mathbf{a}} g)(\mathbf{y})$, which is smaller than $d-1$ iff $T^{-1}\mathbf{a}$ is a fast point for g.

The last part, when $\mathbf{a}$ equals column j of T, follows from $T\mathbf{e}_j = \mathbf{a}$. □

Using Theorem 3, we can transfer the results of Proposition 4 to arbitrary fast points, obtaining thus a general characterisation of fast points. This is different from the characterisation in [3, Theorem 3].

Corollary 2 *Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$. Then f has a fast point iff f can be obtained by an invertible linear change of coordinates from a polynomial function g which, when restricted to the monomials of maximum degree, only depends on $n-1$ (or fewer) variables.*

More precisely, for any $\mathbf{a} \in \mathbb{F}_p^n \setminus \{\mathbf{0}\}$, the following statements are equivalent:

(a) $\mathbf{a}$ *is a fast point for* f

(b) *There is a j for which $a_j \neq 0$ and none of the monomials of maximum total degree (in $y_1, \ldots, y_n$) of $g(y_1, \ldots, y_n) = f(y_1 + a_1 y_j, \ldots, y_{j-1} + a_{j-1} y_j, a_j y_j, y_{j+1} + a_{j+1} y_j, \ldots, y_n + a_n y_j)$ contain y_j.*

(c) *For all j, if $a_j \neq 0$ then none of the monomials of maximum total degree (in $y_1, \ldots, y_n$) of $g(y_1, \ldots, y_n) = f(y_1 + a_1 y_j, \ldots, y_{j-1} + a_{j-1} y_j, a_j y_j, y_{j+1} + a_{j+1} y_j, \ldots, y_n + a_n y_j)$ contain y_j.*

(d) *There is an invertible matrix T which has one column (say column j) equal to $\mathbf{a}$ and none of the monomials of maximum degree in $g(\mathbf{y}) = f(T\mathbf{y})$ contain the variable y_j.*

(e) *For all invertible matrices T which have one column (say column j) equal to $\mathbf{a}$, none of the monomials of maximum degree in $g(\mathbf{y}) = f(T\mathbf{y})$ contain the variable y_j.*

If any of the above equivalent conditions is satisfied then the degree of $(\Delta_{\mathbf{a}} f)(x_1, \ldots, x_n)$ is precisely $d_1 - 1$ where d_1 is the highest total degree among the monomials of $g(y_1, \ldots, y_n)$ that contain y_j.

Example 1 Let us look at the example from [3, Section V], the boolean function $f(x_1, x_2, x_3, x_4) = x_1x_2x_3 + x_1x_2x_4$, which has $(0, 0, 1, 1)$ as a fast point. Although this appears to be a function in 4 variables, writing it as $f(x_1, x_2, x_3, x_4) = x_1x_2(x_3 + x_4)$ it is clear that it is essentially a function in 3 variables, i.e. it can be obtained from the function $g(y_1, y_2, y_3, y_4) = y_1y_2y_3$, which actually only depends on 3 variables, using the invertible linear change of coordinates $y_3 = x_3 + x_4$ and $y_i = x_i$ for $i = 1, 2, 4$.

Next let us look at the example from [6], used also in [4, Section 1], $f(x_1, x_2, x_3, x_4) = x_1x_2x_3 + x_1x_2x_4 + x_2x_3x_4$, which has fast point $(1, 0, 1, 1)$. This looks like a function in 4 variables, and it is not immediately obvious how to write it so that the part of degree 3 depends only on 3 variables. Using the Corollary above and the fact that $(1, 0, 1, 1)$ is a fast point, it turns out that it can be obtained from the polynomial function $g(y_1, y_2, y_3, y_4) = y_1y_2y_3 + y_2y_4$ by the change of coordinates $y_1 = x_1 + x_4$, $y_2 = x_2$, $y_3 = x_3 + x_4$, $y_4 = x_4$. However, if we did not know a fast point for f, finding a suitable change of variable is more difficult, and involves solving a system of linear equations, see below.

Let us consider Corollary 2 in the binary case, and without loss of generality, assume $j = 1$. Determining a change of variables $g(y_1, \ldots, y_n) = f(y_1, y_2 + a_2y_1, \ldots, y_n + a_ny_1)$ such that none of the monomials of maximum total degree (in $y_1, \ldots, y_n$) of g contains y_1 amounts to solving a system of equations in the $n-1$ unknowns $a_2, \ldots, a_n$. The system is obtained by imposing that for each of the $\binom{n-1}{d-1}$ terms of degree d that contain y_1, their coefficient in g is zero. We have thus $\binom{n-1}{d-1}$ equations, and one can verify that in the binary case the equations are linear (Idea of the proof: applying the change of variables to a term, say x_2x_3 produces $(y_2 + a_2y_1)(y_3 + a_3y_1) = y_2y_3 + a_2y_1y_3 + a_3y_1y_2 + a_2a_3y_1^2$, but the last term, whose coefficient is not linear in the a_i, equals actually $a_2a_3y_1$ as a function over $\mathbb{F}_2$, so it is not a term of maximum degree.)

5 Counting functions with fast points

In this section we are interested in counting the functions of degree d, in n variables over $\mathbb{F}_2$ which have fast points. This study was started in [4].

To decide whether a point $\mathbf{a}$ is a fast point for f we only need to look at the monomials of f of degree d, as only they could potentially produce monomials of degree $d-1$ in the derivative. Therefore, instead of counting the functions we will count the equivalence classes of the following equivalence relation on the set of polynomial functions: f is equivalent to g iff $\deg(f) = \deg(g)$ and $\deg(f-g) < \deg(f)$. (In other words two functions are equivalent if when restricted to the monomials of maximum degree, they become equal.) For the rest of this section, whenever we speak of a polynomial function we will mean its equivalence class.

There are $\binom{n}{d}$ terms in n variables which have total degree d and degree at most 1 in each variable. Hence the total number of polynomial functions of degree d over $\mathbb{F}_2$ (in the sense of equivalence classes, as explained above) is $2^{\binom{n}{d}} - 1$.

We saw in Theorem 1 that the fast points form a vector space. We will therefore refine our counting results, by counting the functions that have a space of fast points of a certain dimension.

For any integers n, d, k with $0 \leq d \leq n$ and $0 \leq k \leq n$ we will denote by $F(n,d,k)$ the set of polynomial functions over $\mathbb{F}_2$ (equivalence classes in the sense above) in n variables, of degree d and with a space of fast points of dimension k. Since these sets form a partition of the set of polynomial functions of degree d, we have

$$\sum_{k=0}^{n} |F(n,d,k)| = 2^{\binom{n}{d}} - 1. \tag{1}$$

Like in [4] we denote by $F(n,d)$ the set of polynomials of degree d in n variables that have (any number of) non-trivial fast points. Hence:

$$|F(n,d)| = \sum_{k=1}^{n} |F(n,d,k)|. \tag{2}$$

Using equation (1) this can be rewritten as:

$$|F(n,d)| = 2^{\binom{n}{d}} - 1 - |F(n,d,0)|. \tag{3}$$

For a start we examine the set of functions that have a given set of canonical basis vectors as fast points:

Lemma 1

(i) *The set of polynomial functions over $\mathbb{F}_2$ in n variables $x_1, \ldots, x_n$ and of degree d which have $\langle \mathbf{e}_{n-k+1}, \ldots, \mathbf{e}_n \rangle$ included in their space of fast points is precisely the set of polynomial functions over $\mathbb{F}_2$ in $n-k$ variables $x_1, \ldots, x_{n-k}$ and of degree d. (Note this set is empty if $d > n-k$.) The cardinality of this set is therefore $2^{\binom{n-k}{d}} - 1$.*

(ii) *The set of polynomial functions over $\mathbb{F}_2$ in n variables $x_1, \ldots, x_n$ and of degree d which have their space of fast points equal to $\langle \mathbf{e}_{n-k+1}, \ldots, \mathbf{e}_n \rangle$ is equal to the set $F(n-k,d,0)$ (when considered as functions in n variables).*

Proof The proof of (i) follows from Proposition 4.

For (ii) we prove first the "$\subseteq$" part. Assume f has the space of fast points equal to $\langle \mathbf{e}_{n-k+1}, \ldots, \mathbf{e}_n \rangle$. By (i), f is a function in $x_1, \ldots, x_{n-k}$. Moreover, we show that $f \in F(n-k, d, 0)$, i.e. f has no fast point when considered as a function in $x_1, \ldots, x_{n-k}$. Assume, for a contradiction, that f has such a non-trivial fast point $\mathbf{a} = (a_1, \ldots, a_{n-k}) \neq \mathbf{0}$. But then the n-tuple $(a_1, \ldots, a_{n-k}, 0, \ldots, 0)$ is also a non-trivial fast point of f as function over $x_1, \ldots, x_n$. However, this fast point is not in the space $\langle \mathbf{e}_{n-k+1}, \ldots, \mathbf{e}_n \rangle$. Contradiction.

For the reverse inclusion, "$\supseteq$", let $f \in F(n-k, d, 0)$. As a function in $x_1, \ldots, x_n$, f has a space of fast points that contains $\langle \mathbf{e}_{n-k+1}, \ldots, \mathbf{e}_n \rangle$, by (i). We show that f has no other fast points. Assume $\mathbf{a} = (a_1, \ldots, a_n) \neq \mathbf{0}$ is a fast point, and denote by $\mathbf{b} = (a_1, \ldots, a_{n-k})$. Since f does not depend on $x_{n-k+1}, \ldots, x_n$, we have $\Delta_{\mathbf{a}} f = \Delta_{\mathbf{b}} f$, so $\mathbf{b}$ is a fast point of f as a function in $n-k$ variables. Since $f \in F(n-k, d, 0)$, this means $\mathbf{b} = \mathbf{0}$, so $\mathbf{a} \in \langle \mathbf{e}_{n-k+1}, \ldots, \mathbf{e}_n \rangle$. □

The key to computing $|F(n, d, k)|$ will be by doing a convenient change of variables.

Theorem 4 *Let $f : K^n \to K$ be a polynomial function. The space of fast points of f has basis $\mathbf{a}_1, \ldots, \mathbf{a}_k$ iff the space of fast points of g has basis $\mathbf{e}_{n-k+1}, \ldots, \mathbf{e}_n$, where $g(\mathbf{y}) = f(T\mathbf{y})$ and T is any invertible matrix having the last k columns equal to $\mathbf{a}_1, \ldots, \mathbf{a}_k$.*

Proof Using Theorem 3, we know that $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are all fast points for f iff $\mathbf{e}_{n-k+1}, \ldots, \mathbf{e}_n$ are all fast points for g. Moreover any point $\mathbf{a}$ is a fast point for f iff $T^{-1}\mathbf{a}$ is a fast point for g. On the other hand, simple linear algebra yields $\mathbf{a} \in \langle \mathbf{a}_1, \ldots, \mathbf{a}_k \rangle$ iff $T^{-1}\mathbf{a} \in \langle T^{-1}\mathbf{a}_1, \ldots, T^{-1}\mathbf{a}_k \rangle = \langle \mathbf{e}_{n-k+1}, \ldots, \mathbf{e}_n \rangle$. □

We can now use Theorem 4 to generalise Lemma 1 to an arbitrary space of fast points:

Lemma 2 *Let V be a vector subspace of $\mathbb{F}_2^n$ of dimension k. Let T be any $n \times n$ invertible matrix whose last k columns are a basis for V.*

(i) *The set of polynomial functions over $\mathbb{F}_2$ in n variables of degree d which have V included in their space of fast points equals the set of polynomial functions of the form $g(T^{-1}\mathbf{x})$ where g ranges over all polynomial functions over $\mathbb{F}_2$ in $n-k$ variables $x_1, \ldots, x_{n-k}$ and of degree d. This set has cardinality $2^{\binom{n-k}{d}} - 1$ for $0 \leq d \leq n-k$ and is empty otherwise.*

(ii) *The set of polynomial functions over $\mathbb{F}_2$ in n variables of degree d which have their set of fast points equal to V equals*

$$\{g(T^{-1}\mathbf{x}) | g \in F(n-k, d, 0)\}$$

Proof Use Lemma 1 and Theorem 4. □

Lemma 2(i) above also gives an alternative proof of the following result:

Proposition 5 [4, Theorem 3.1] *The space of fast points of a polynomial function f over $\mathbb{F}_2$ has dimension at most $n - \deg(f)$.*

Hence $F(n, d, k) = \emptyset$ for $k > n - d$. In equations (1) and (2), in the upper bound of the summation we can replace n by $n - d$.

The first major step in our counting problem will be to construct the set $F(n,d,k)$ from the set $F(n-k,d,0)$, thus reducing the computation of $|F(n,d,k)|$ to the computation of $|F(n-k,d,0)|$:

Theorem 5

(i) *We have*
$$F(n,d,k) = \bigcup_{T} \{g(T\mathbf{x}) | g \in F(n-k,d,0)\}$$
with T ranging over all the invertible $n \times n$ matrices.

(ii) *Denote by V_i, with $i = 1, \ldots, \begin{pmatrix} n \\ k \end{pmatrix}_2$ all the spaces of $\mathbb{F}_2^n$ of dimension k. For each such vector space V_i consider an invertible $n \times n$ matrix T_i such that the last k columns of T_i generate V_i. Then*
$$F(n,d,k) = \bigcup_{i=1}^{\begin{pmatrix} n \\ d \end{pmatrix}_2} \{g\left(T_i^{-1}\mathbf{x}\right) | g \in F(n-k,d,0)\}$$
and the sets in the union above are disjoint.

(iii) *We have*
$$|F(n,d,k)| = \begin{pmatrix} n \\ k \end{pmatrix}_2 |F(n-k,d,0)|.$$

Proof

(i) Let $g \in F(n-k,d,0)$. By Lemma 1(ii), g, when viewed as a function in n variables, has a space of fast points of dimension k. By Theorem 4, $g(T\mathbf{x})$ also has a space of fast points of dimension k, so it is in $F(n,d,k)$. The reverse follows from (ii).

(ii) We can partition $F(n,d,k)$ into the following disjoint sets:
$$F(n,d,k) = \bigcup_{V_i} \{f | \deg(f) = d, \text{ the space of fast points of } f \text{ is } V_i\}.$$
Using Lemma 2(ii), we have that
$$\{f | \deg(f) = d, \text{ the space of fast points of } f \text{ is } V_i\} = \{g\left(T_i^{-1}\mathbf{x}\right) | g \in F(n-k,d,0)\}.$$

(iii) follows from (ii) immediately.

□

Using the Theorem above and equation (2) we obtain:
$$|F(n,d)| = \sum_{k=1}^{n-d} \begin{pmatrix} n \\ k \end{pmatrix}_2 |F(n-k,d,0)|$$

Using equation (3) this can be rewritten to obtain a linear recurrence relation on n for $|F(n,d)|$:
$$|F(n,d)| = \sum_{k=1}^{n-d} \begin{pmatrix} n \\ k \end{pmatrix}_2 \left(2^{\begin{pmatrix} n-k \\ d \end{pmatrix}} - 1 - |F(n-k,d)|\right) \quad (4)$$

Using the initial conditions $|F(d,d)| = 0$ (there is only one term of degree d in d variables, and it has no non-trivial fast points) we can already compute recursively any $|F(n,d)|$ (and then any $|F(n,d,k)|$).

For a few particular cases we can immediately obtain an explicit formula:

Proposition 6

(i) $|F(n,d,n-d)| = \binom{n}{d}_2$.
(ii) $|F(n,d,n-d-1)| = 0$.
(iii) *For degree* $d=1$ *we have* $|F(n,1)| = |F(n,1,n-1)| = 2^n-1$ *and* $|F(n,1,k)| = 0$ *for all* $0 \le k < n-1$. *In other words, all functions of degree one have a space of fast points of dimension* $n-1$.
(iv) *For degree* $d=n-1$ *we have* $|F(n,n-1)| = |F(n,n-1,1)| = 2^n-1$, *and* $|F(n,n-1,0)| = 0$. *In other words all functions of degree* $n-1$ *have exactly one non-trivial fast point.*
(v) *For degree* $d=n-2$ *we have* $|F(n,n-2)| = |F(n,n-2,2)| = (2^n-1)(2^{n-1}-1)/3$. *This implies that if a function of degree* $n-2$ *has non-trivial fast points, then it has exactly 3 such points (to form, together with* $\mathbf{0}$, *a space of dimension 2).*

Proof First note that, as mentioned above there is only one term of degree n, and it has no non-trivial fast points, hence $|F(n,n,0)| = 1$ for all n. For (i), using Theorem 5 we have:

$$|F(n,d,n-d)| = \binom{n}{n-d}_2 |F(d,d,0)| = \binom{n}{n-d}_2 = \binom{n}{d}_2$$

Using (i) and equation (1) we prove (ii), first for $d=n-1$ and then for arbitrary d:

$$|F(n,n-1,0)| = 2^{\binom{n}{n-1}} - 1 - |F(n,n-1,1)| = 2^n - 1 - \binom{n}{n-1}_2 = 0$$

$$|F(n,d,n-d-1)| = \binom{n}{k}_2 |F(d+1,d,0)| = 0$$

Using (i) and (ii) and equation (1) we can then easily obtain (iii)-(v). □

In Duan et al., a formula for $|F(n,d)|$ was only obtained for the cases of degree $d=n-1$ and $d=n-2$, see [4, Theorem 3.3, Theorem 3.6]. Proposition 6 (iv) and (v) above gives alternative proofs of their results.

For the rest of the cases in [4], some experimental results were computed by enumerating each single polynomial function and checking whether it has a fast point. This exhaustive approach has a very high computational cost (higher than exponential in n) and therefore already for $n=7$ some entries were left blank in their tables (namely $n=7, k=3,4$ and $n=8, k=3,4,5$). We can fill in these missing values at a very modest computational cost (the complexity is polynomial in n), using the recurrence (4). Our data for values of n up to 8 is presented in Section 7.

In addition to the recurrence relation (4), we also obtained an explicit formula for $|F(n,d)|$:

Theorem 6

$$|F(n,d)| = \sum_{i=1}^{n-d} (-1)^{i-1} 2^{\frac{i(i-1)}{2}} \binom{n}{i}_2 \left(2^{\binom{n-i}{d}} - 1\right)$$

Proof For any vector subspace $V \subseteq \mathbb{F}_2^n$, denote by A_V the set of functions in n variables, of degree d, which have V included in their space of fast points. We have

$$F(n,d) = \bigcup_{\mathbf{a} \in \mathbb{F}_2^n \setminus \{0\}} A_{\langle \mathbf{a} \rangle}.$$

To compute $|F(n,d)|$ we employ an inclusion-exclusion formula and the fact that $A_{V_1} \cap A_{V_2} = A_{\langle V_1 \cup V_2 \rangle}$:

$$|F(n,d)| = \sum_{\mathbf{a}_1 \in \mathbb{F}_2^n \setminus \{0\}} |A_{\langle \mathbf{a}_1 \rangle}| - \sum_{\mathbf{a}_1, \mathbf{a}_2} |A_{\langle \mathbf{a}_1, \mathbf{a}_2 \rangle}| + \sum_{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3} |A_{\langle \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3 \rangle}| - \dots$$

Obviously, the sets above are not all distinct. Namely for each vector space V, $|A_V|$ appears in the formula above a number of times equal to the number of spanning sets of V. More precisely, $|A_V|$ is added once for each spanning set of V of odd cardinality and subtracted once for each spanning set of V of even cardinality. So overall $|A_V|$ is counted a number of times equal to the number of odd cardinality spanning sets of V minus the number of even cardinality spanning sets of V. By the technical Lemma 3 in the Appendix, this number is equal to

$$(-1)^{i-1} 2^{\frac{i(i-1)}{2}}.$$

where $i = \dim(V)$. By Lemma 2(i), $|A_V| = 2^{\binom{n-i}{d}} - 1$. The maximum dimension of the space of fast points of a function in $F(n,d)$ is $n-d$. For each dimension i between 1 and $n-d$ there are $\binom{n}{i}_2$ subspaces V of dimension i. Putting everything together in the inclusion-exclusion formula above we obtain

$$|F(n,d)| = \sum_{i=1}^{n-d} (-1)^{i-1} 2^{\frac{i(i-1)}{2}} \binom{n}{i}_2 \left(2^{\binom{n-i}{d}} - 1 \right).$$

□

Using Theorems 5 (iii) and 6 and equation (3) we can now obtain explicit formulae for all $|F(n,d,k)|$:

Corollary 3

$$|F(n,d,k)| = \binom{n}{k}_2 \sum_{i=0}^{n-k-d} (-1)^i \, 2^{\frac{i(i-1)}{2}} \binom{n-k}{i}_2 \left(2^{\binom{n-k-i}{d}} - 1 \right)$$

Remark 2 It might be possible to prove Theorem 6 by using some method of solving the recurrence (4), or by induction, using again the recurrence (4). We did not find any simple proof along those lines, so we preferred the current proof that uses a natural inclusion-exclusion argument.

6 Higher order differentiation and change of variables

In this section we give some results regarding the use of change of variables with higher order differentiation.

6.1 Background

We recall in this subsection a few known or straightforward results. An explicit formula for higher order derivatives can be obtained easily by induction:

Proposition 7 *Let $f : K^n \to K$ be a function in n variables $x_1, \ldots, x_n$. Let $\mathbf{a}_1, \ldots, \mathbf{a}_k \in K^n$ not necessarily distinct. Then*

$$\Delta^{(k)}_{\mathbf{a}_1,\ldots,\mathbf{a}_k} f(\mathbf{x}) = \sum_{\mathbf{u}=(u_1,\ldots,u_k)\in\{0,1\}^k} (-1)^{k-\mathrm{w}(\mathbf{u})} f(\mathbf{x} + u_1\mathbf{a}_1 + \cdots + u_k\mathbf{a}_k)$$

where w() *denotes the Hamming weight.*

When K is the binary field $\mathbb{F}_2$, in the last formula above the summation is over the elements of the vector space generated by $\mathbf{a}_1, \ldots, \mathbf{a}_k$. We have therefore:

Corollary 4 *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a function in n variables $x_1, \ldots, x_n$. Let $\mathbf{a}_1, \ldots, \mathbf{a}_k \in \mathbb{F}_2^n$ be linearly independent and let V be the vector space generated by $\mathbf{a}_1, \ldots, \mathbf{a}_k$. Then*

$$\Delta^{(k)}_{\mathbf{a}_1,\ldots,\mathbf{a}_k} f(\mathbf{x}) = \sum_{\mathbf{v}\in V} f(\mathbf{x} + \mathbf{v}).$$

Corollary 5 *Let $\mathbf{a}_1, \ldots, \mathbf{a}_k \in \mathbb{F}_2^n$ and $\mathbf{b}_1, \ldots, \mathbf{b}_k \in \mathbb{F}_2^n$ be two sets of linearly independent vectors. We have:*

$\Delta^{(k)}_{\mathbf{a}_1,\ldots,\mathbf{a}_k} = \Delta^{(k)}_{\mathbf{b}_1,\ldots,\mathbf{b}_k}$
iff
$\mathbf{a}_1, \ldots, \mathbf{a}_k$ and $\mathbf{b}_1, \ldots, \mathbf{b}_k$ generate the same vector space.

Depending on the values of the $\mathbf{a}_1, \ldots, \mathbf{a}_k$ and the characteristic of the field, $\Delta^{(k)}_{\mathbf{a}_1,\ldots,\mathbf{a}_k} f$ could collapse, becoming the identical zero function regardless of the function f. This happens, for example, if the characteristic is 2 and $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are not linearly independent. Hence in $\mathbb{F}_2$ we will always assume that $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are linearly independent.

6.2 Higher order differentiation and generalisation of fast points

Theorem 2 can be generalised as follows:

Theorem 7 *Let $\mathbf{a}_1, \ldots, \mathbf{a}_k \in \mathbb{F}_2^n$ be linearly independent. Let T be an invertible $n \times n$ matrix over K. Let f be a function of n variables $f : K^n \to K$; denote by g the function obtained from f via the change of variables defined by T, namely: $g(\mathbf{y}) = f(T\mathbf{y})$. Then we have:*

$$\left(\Delta^{(k)}_{\mathbf{a}_1,\ldots,\mathbf{a}_k} f\right)(T\mathbf{y}) = \left(\Delta^{(k)}_{T^{-1}\mathbf{a}_1,\ldots,T^{-1}\mathbf{a}_k} g\right)(\mathbf{y}).$$

In particular if there exist k columns of T, say columns $i_1, \ldots, i_k$ which equal $\mathbf{a}_1, \ldots, \mathbf{a}_k$ (in the case of $K = \mathbb{F}_2$, it suffices that there exist k columns of T that generate the vector space $\langle \mathbf{a}_1, \ldots, \mathbf{a}_k \rangle$) then we have:

$$\left(\Delta^{(k)}_{\mathbf{a}_1,\ldots,\mathbf{a}_k} f\right)(T\mathbf{y}) = \left(\Delta^{(k)}_{\mathbf{e}_{i_1},\ldots,\mathbf{e}_{i_k}} g\right)(\mathbf{y}).$$

Proof This Theorem can be proven by repeated application of Theorem 2, or directly by applying the change of variables $\mathbf{x} = T\mathbf{y}$ in the formula given in Proposition 7:

$$\begin{aligned}\left(\Delta^{(k)}_{\mathbf{a}_1,\ldots,\mathbf{a}_k} f\right)(T\mathbf{y} &= \sum_{\mathbf{u}\in\{0,1\}^k} f(T\mathbf{y} + u_1\mathbf{a}_1 + \cdots + u_k\mathbf{a}_k)\\ &= \sum_{\mathbf{u}\in\{0,1\}^k} f(T(\mathbf{y} + u_1T^{-1}\mathbf{a}_1 + \cdots + u_kT^{-1}\mathbf{a}_k))\\ &= \sum_{\mathbf{u}\in\{0,1\}^k} g(\mathbf{y} + u_1T^{-1}\mathbf{a}_1 + \cdots + u_kT^{-1}\mathbf{a}_k)\\ &= \left(\Delta^{(k)}_{T^{-1}\mathbf{a}_1,\ldots,T^{-1}\mathbf{a}_k} g\right)(\mathbf{y}).\end{aligned}$$

□

An equivalent formulation of the Theorem above is to say that the following diagram commutes:

$$\begin{array}{ccc}\mathcal{F}(K^n, K) & \xrightarrow{\Delta^{(k)}_{\mathbf{a}_1,\ldots,\mathbf{a}_k}} & \mathcal{F}(K^n, K)\\ \Phi_T \downarrow & & \downarrow \Phi_T\\ \mathcal{F}(K^n, K) & \xrightarrow{\Delta^{(k)}_{\mathbf{e}_{i_1},\ldots,\mathbf{e}_{i_k}}} & \mathcal{F}(K^n, K)\end{array}$$

with the notations defined after Theorem 2.

We introduce a generalisation of the notion of fast point, called a fast space (not to be confused with the space of fast points).

Definition 5 Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$. A vector subspace $V \subseteq \mathbb{F}_2^n$ is called a *fast space* for f if there is a basis $\mathbf{a}_1, \ldots, \mathbf{a}_k$ of V such that $\deg\left(\Delta^{(k)}_{\mathbf{a}_1,\ldots,\mathbf{a}_k} f\right) < \deg(f) - k$.

Note that Corollary 5 ensures that if one basis of V satisfies the condition in the definition above, then all the bases do.

It is easy to see that if a vector space V contains a fast point, then it is a fast space. However the reverse is not always true: a vector space V can be a fast space but not contain any fast points. The following example illustrates this situation:

Example 2 Let $f(x_1, x_2, x_3, x_4, x_5) = x_1x_3x_4 + x_2x_3x_5 + x_1x_2$. The space $V = \langle \mathbf{e}_1, \mathbf{e}_2\rangle$ is a fast space, as $\Delta^{(2)}_{\mathbf{e}_1,\mathbf{e}_2} f = 1$ has degree $0 < \deg(f) - 2 = 1$. However one can check that none of the elements of V, namely $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_1 + \mathbf{e}_2$ is a fast point as differentiating with respect to any of them produces a polynomial of degree two: $\Delta_{\mathbf{e}_1} f = x_3x_4 + x_2$, $\Delta_{\mathbf{e}_2} f = x_3x_5 + x_1$, $\Delta_{\mathbf{e}_1+\mathbf{e}_2} f = x_3x_4 + x_3x_5 + x_1 + x_2 + 1$.

We can extend Proposition 4 to higher order derivatives:

Proposition 8 *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a polynomial function and $i_1, \ldots, i_k$ be distinct indices in $\{1, \ldots, n\}$. The total degree of $\Delta^{(k)}_{\mathbf{e}_{i_1},\ldots,\mathbf{e}_{i_k}} f$ equals $d_{i_1,\ldots,i_k} - k$ where $d_{i_1,\ldots,i_k}$ is the highest total degree among the monomials of f that are divisible by the term $x_{i_1} \ldots x_{i_k}$.*

In particular, $\langle \mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_k}\rangle$ is a fast space for f iff none of the terms of highest degree in f are divisible by the term $x_{i_1} \ldots x_{i_k}$.

Proof Denote $t = x_{i_1} \dots x_{i_k}$ and factor out the term t, writing $f = tf_1 + f_2$ with f_1 not depending on any of the variables $x_{i_1}, \dots, x_{i_k}$ and none of the terms of f_2 divisible by t. Using [2, Theorem 1] and [3, Section IV B] or [5, Section 4.1] we have that $\Delta^{(k)}_{\mathbf{e}_{i_1},\dots,\mathbf{e}_{i_k}} f = f_1$. □

Using this result and Theorem 7 we can give a characterisation of fast spaces:

Theorem 8 *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a polynomial function and $\mathbf{a}_1, \dots, \mathbf{a}_k \in \mathbb{F}_2^n$ linearly independent. Let T be an invertible matrix constructed as follows: the last k columns are $\mathbf{a}_1, \dots, \mathbf{a}_k$ and the remaining $n-k$ columns are any vectors so that T is invertible. Consider the change of variables $\mathbf{x} = T\mathbf{y}$.*

The total degree of $\Delta^{(k)}_{\mathbf{a}_1,\dots,\mathbf{a}_k} f$ is $\deg(f) - k - k'$ iff in $f(T\mathbf{y})$ the highest total degree among the monomials divisible by the term $y_{n-k+1} \cdots y_n$ is $\deg(f) - k'$.

In particular, $\langle \mathbf{a}_1, \dots, \mathbf{a}_k \rangle$ is a fast space for f iff none of the terms of highest degree in $f(T\mathbf{y})$ are divisible by the term $y_{n-k+1} \cdots y_n$.

One could also attempt to count functions that admit fast spaces, but the analysis becomes quite difficult and will be left as a topic of possible future work.

7 Numerical results

Using the recurrence relation (4), or alternatively the explicit formulae given in Theorem 6 and Corollary 3, we computed numerical values for polynomials with fast points in up to 8 variables and presented them in Tables 1, 2, 3, 4, 5 and 6 (Tables 3–6 are in the Appendix) with each individual table corresponding to a given number of variables. All the numbers refer to equivalence classes, as explained at the beginning of Section 5.

For each number of variables n, the main part of each table displays $|F(n, d, k)|$ (the number of polynomial functions of degree d with a space of fast points of dimension k) in row d, column k.

The last two columns in each table give in row d the value of $|F(n, d)|$ (the number of polynomial functions of degree d with non-trivial fast points) in absolute terms, and also as a proportion of the total number $\left(2^{\binom{n}{d}} - 1\right)$ of polynomials of degree d. For $n = 7$ and $n = 8$, the values of $|F(n, d)|$ and of $2^{\binom{n}{d}} - 1$ are also plotted in Fig. 1 in the Appendix. Since the former is much lower than the latter, we used a logarithmic ($\log_2$) scale.

Table 1 Number of functions with fast points in 7 variables

$d\backslash k$	0	1	2	3	4	5	6	$\lvert F(7,d)\rvert$	Ratio
1	0	0	0	0	0	0	127	127	1
2	0	1763776	0	330708	0	2667		2097151	1
3	34231364608	126046992	2314956	0	11811			128373759	0.00374
4	34355647824	4078732	0	11811				4090543	0.00012
5	2094484	0	2667					2667	0.00127
6	0	127						127	1
7	1							0	0

Table 2 Number of functions with fast points in 8 variables

$d\backslash k$	0	1	2	3	4	5	6	7	$\lvert F(8,d)\rvert$	Ratio
1	0	0	0	0	0	0	0	255	255	1
2	112881664	0	149920960	0	5622036	0	10795		155553791	0.57948
3	7.20489E+16	8.729E+12	10713994320	84330540	0	97155			8.7398E+12	0.00012
4	1.18059E+21	8.76069E+12	346692220	0	200787				8.76104E+12	7.42089E-09
5	7.20576E+16	534093420	0	97155					534190575	7.41338E-09
6	268424660	0	10795						10795	0.00004
7	0	255							255	1
8	1								0	0

The values of $|F(n,d)|$ (in absolute terms, and also as a proportion of the total number of polynomials of degree d) were given in [4, Table A.1] for n up to 8, except for degrees 3 and 4 in 7 variables and for degrees 3, 4, and 5 in 8 variables, which could not be computed due to the very high computational complexity.

Our results confirm their existing results and also fill in the missing entries in their table, all at negligible computational cost (less than 1 second for each n, even on a low specification computer).

Looking at the tables we can notice some trends. For $3 \leq d \leq n-2$, the ratio is very small, i.e. a very small proportion of functions have fast points. For each fixed degree d, (again, with $3 \leq d \leq n-2$), the number of functions decreases as k increases (excluding the last two entries, $k = n-d-1$ and $k = n-d$ for which results are given in Proposition 6). So most of the functions have no fast points, much fewer functions have one fast point, even fewer functions have 3 fast points and so on. Asymptotic results will be obtained in the next section.

8 Cryptographic consequences

Throughout this section we only work with functions over $\mathbb{F}_2$.

We are given a cryptographic function f as a "black box" function with n input bits and one output bit. In order to cryptanalyse the function, we differentiate f several times. Note that $\Delta^{(k)}_{\mathbf{a}_1,\ldots,\mathbf{a}_k} f$ is also a "black box": its output can be evaluated for any given input, by doing 2^k calls to f (see Proposition 7). The attacks that use this approach hope to determine some "non-random" properties of $\Delta^{(k)}_{\mathbf{a}_1,\ldots,\mathbf{a}_k} f$.

The first consequence of our previous analysis (Theorem 7) is that instead of computing $\Delta^{(k)}_{\mathbf{a}_1,\ldots,\mathbf{a}_k} f$ we can first replace f by another "black box" function g, which simply feeds $T\mathbf{x}$ into f, where T is any matrix whose first k columns are $\mathbf{a}_1,\ldots,\mathbf{a}_k$ (or any other k columns that generate the same vector space). We then compute $\Delta^{(k)}_{\mathbf{e}_1,\ldots,\mathbf{e}_k} g$. These two methods are equivalent (in the sense that the degree of the resulting function is the same), but depending on the particular application, there may be advantages of doing one or the other.

Having a low degree (preferably degree one) is a particularly useful property of $\Delta^{(k)}_{\mathbf{a}_1,\ldots,\mathbf{a}_k} f$ and this is the property exploited by the AIDA/cube attacks. Since evaluating this function takes 2^k calls to f, we hope that a low degree is reached for a relatively low k (lower than $\deg(f)-1$). It is therefore useful for an attacker if $\mathbf{a}_1$ is a fast point for f (and $\mathbf{a}_2$ is a fast point for $\Delta_{\mathbf{a}_1} f$ etc.) So let us concentrate on one differentiation.

Assume that f is random, of degree d, in the sense that it is picked out of a uniform distribution over the set of polynomials functions of degree d in n variables. Throughout the

rest of this section we assume that $3 \leq d \leq n-3$. For $d = 1, n-2, n-1$ see Proposition 6 or [4, Theorems 3.3 and 3.6] and for $d = 2$ see [4, Section 3.4].

We then pick a point $\mathbf{a} \in \mathbb{F}_2^n \setminus \{\mathbf{0}\}$ and compute the derivative of f w.r.t. $\mathbf{a}$. We pick $\mathbf{a}$ assuming a uniform distribution on $\mathbb{F}_2^n \setminus \{\mathbf{0}\}$, and independently from f.

The first question we can ask is: what is the probability of $\mathbf{a}$ being a fast point for f?

Proposition 9 *The probability of* $\mathbf{a}$ *being a fast point for* f *only depends on the degree* d *of* f *and is given by*

$$\frac{2^{\binom{n-1}{d}} - 1}{2^{\binom{n}{d}} - 1} \approx \frac{1}{2^{\binom{n-1}{d-1}}}$$

Proof Apply Lemma 2 (i) for $V = \langle \mathbf{a} \rangle$. Since $\dim(V) = 1$, there are $2^{\binom{n-1}{d}} - 1$ functions that have the fast point $\mathbf{a}$. For the approximation we have

$$\frac{2^{\binom{n-1}{d}} - 1}{2^{\binom{n}{d}} - 1} \approx \frac{2^{\binom{n-1}{d}}}{2^{\binom{n}{d}}} = \frac{1}{2^{\binom{n}{d} - \binom{n-1}{d}}} = \frac{1}{2^{\binom{n-1}{d-1}}}.$$

□

Consequently the strongest functions (i.e. least likely to have fast points) are those of degree close to half the number of variables. The probability of having a fast point is lowest when the degree is close to half the number of variables, and increases as we move away towards higher degrees; it also increases as we move from $n/2$ towards lower degrees. However in all cases the probability of $\mathbf{a}$ being a fast point is extremely small, and it tends to zero as n goes to infinity.

The second question we may ask is, given a random f as above, what is the probability that f has fast points? Again this only depends on the degree d of f, and is given by

$$\frac{|F(n,d)|}{2^{\binom{n}{d}} - 1}.$$

We will now estimate this quantity. These approximations are valid when d and $n-k-d$ are both greater than 2.

In the sum from Theorem 6:

$$|F(n,d)| = \sum_{i=1}^{n-d} (-1)^{i-1} \, 2^{\frac{i(i-1)}{2}} \begin{bmatrix} n \\ i \end{bmatrix}_2 (2^{\binom{n-i}{d}} - 1)$$

we compare the absolute value of successive terms by estimating the ratio of term $i+1$ and term i:

$$\frac{2^{\frac{i(i+1)}{2}} \begin{bmatrix} n \\ i+1 \end{bmatrix}_2 \left(2^{\binom{n-i-1}{d}} - 1\right)}{2^{\frac{i(i-1)}{2}} \begin{bmatrix} n \\ i \end{bmatrix}_2 \left(2^{\binom{n-i}{d}} - 1\right)} = 2^i \cdot \frac{2^{n-i} - 1}{2^{i+1} - 1} \cdot \frac{\left(2^{\binom{n-i-1}{d}} - 1\right)}{\left(2^{\binom{n-i}{d}} - 1\right)}$$

$$\approx \frac{2^{n-i-1}}{2^{\binom{n-i}{d} - \binom{n-i-1}{d}}} \approx \frac{1}{2^{\binom{n-i-1}{d-1} - n + i + 1}}$$

Hence the summands decrease rapidly in absolute value as i increases, and since they have alternating signs we can approximate the sum by the first term (which also gives an upper bound):

$$|F(n,d)| \approx \binom{n}{1}_2 (2^{\binom{n-1}{d}} - 1) \approx 2^{n+\binom{n-1}{d}}.$$

Hence

$$\frac{|F(n,d)|}{2^{\binom{n}{d}} - 1} \approx \frac{2^{n+\binom{n-1}{d}}}{2^{\binom{n}{d}}} = \frac{1}{2^{\binom{n-1}{d-1}-n}}.$$

Therefore this ratio tends to zero as n goes to infinity as long as d stays in the specified range ($3 \leq d \leq n-3$). More precisely, for any sequence $(d_n)_{n\in\mathbb{N}}$ with $3 \leq d_n \leq n-3$ we have

$$\lim_{n\to\infty} \frac{|F(n,d_n)|}{2^{\binom{n}{d_n}} - 1} = 0.$$

The decrease is very rapid, so the ratio is already very small for all n that are of interest in cryptographical applications. In [4, Remark 12] it is conjectured that this ratio decreases like a geometric series in n. We see here that in fact the decrease is much more rapid than that.

We can also see how the ratio of the logarithms behaves:

$$\frac{\log_2(|F(n,d)|)}{\log_2(2^{\binom{n}{d}} - 1)} \approx \frac{\binom{n-1}{d}}{\binom{n}{d}} = \frac{n-d}{n}$$

For example, if $d = n/2$, the above result tells us that the number of functions with fast points is roughly the square root of the total number of functions.

However, the fact that f has a fast point does not mean it is easy to find that fast point. For a given "black box" function f, the larger its space of fast points, the better the chances of an arbitrarily chosen point to be a fast point. So we can refine the previous question as: what is the probability that f has "a lot" of fast points? What is the probability that f has "very few" fast points?

These probabilities are:

$$\frac{|F(n,d,k)|}{2^{\binom{n}{d}} - 1}$$

with k close to $n-d$ for "lots" of fast points, and k small for "very few" fast points.

Using the the same arguments as above, for $3 \leq d \leq n-k-3$ we can approximate $|F(n,d,k)|$ by the first term in the sum in Corollary 3, i.e.

$$|F(n,d,k)| \approx \binom{n}{k}_2 \left(2^{\binom{n-k}{d}} - 1\right) \approx 2^{k(n-k)+\binom{n-k}{d}}$$

We have:

$$\frac{|F(n,d,k+1)|}{|F(n,d,k)|} \approx \frac{1}{2^{\binom{n-k-1}{d-1}-n+2k+1}}$$

hence

$$|F(n,d,k)| \gg |F(n,d,k+1)|$$

for every k. This means that lots of functions have no fast points, much fewer have 1 fast point, even fewer have 3 fast points etc.

Approximating the Gaussian binomial coefficients as follows:

$$\binom{n}{k}_2 = \frac{\prod_{i=n-k+1}^{n}(2^i-1)}{\prod_{i=1}^{k}(2^i-1)} \approx \frac{\prod_{i=n-k+1}^{n}2^i}{\prod_{i=1}^{k}2^i} = 2^{k(n-k)}$$

we have:

$$\frac{\log_2(|F(n,d,k+1)|)}{\log_2(|F(n,d,k)|)} \approx \frac{\binom{n-k-1}{d}}{\binom{n-k}{d}} = \frac{n-k-d}{n}.$$

Finally, let us also look at our results from the point of view of the designer of a cryptographic function, rather than the attacker. We should avoid, as much as possible, that our function has fast points. In light of our characterisation in Corollary 2, this means that we should avoid functions in which some of the variables do not appear in any of the terms of maximum total degree. Moreover, we should avoid functions that superficially look like they depend on all the variables, but with a suitable change of variables, this is no longer the case (see Example 1).

9 Conclusion

Using linear changes of variable we obtained an alternative characterisation of the "fast points" introduced by Duan and Lai in [3]. We also completed the counting of functions with fast points commenced by Duan et al. in [4], giving explicit formulae in the general case. We discussed the cryptographic significance of our results. Fast points have very low probability to be found if the function and the candidate fast point are chosen uniformly at random and independently of each other. An attacker needs therefore to try to exploit extra knowledge about the function in order to increase their chances of finding fast points.

Appendix

Lemma 3 *Let $V \subseteq \mathbb{F}_p^n$ be a vector space of dimension k. Let $S_0(k)$ denote the set of spanning sets of V with even cardinality (we do not allow $\mathbf{0}$ to be an element of a spanning set). Similarly $S_1(k)$ for odd cardinality. Then*

$$|S_1(k)| - |S_0(k)| = (-1)^{k-1} p^{\frac{k(k-1)}{2}}$$

Table 3 Number of functions with fast points in 3 variables

$d\backslash k$	0	1	2	$\lvert F(3,d)\rvert$	Ratio
1	0	0	7	7	1
2	0	7		7	1
3	1			0	1

Table 4 Number of functions with fast points in 4 variables

$d\backslash k$	0	1	2	3	$\lvert F(4,d)\rvert$	Ratio
1	0	0	0	15	15	1
2	28	0	35		35	0.55556
3	0	15			15	1
4	1				0	0

Table 5 Number of functions with fast points in 5 variables

$d\backslash k$	0	1	2	3	4	$\lvert F(5,d)\rvert$	Ratio
1	0	0	0	0	31	31	1
2	0	868	0	155		1023	1
3	868	0	155			155	0.15152
4	0	31				31	1
5	1					0	0

Table 6 Number of functions with fast points in 6 variables

$d\backslash k$	0	1	2	3	4	5	$\lvert F(6,d)\rvert$	Ratio
1	0	0	0	0	0	63	63	1
2	13,888	0	18,228	0	651		18,879	0.57616
3	992,496	54,684	0	1,395			56,079	0.05348
4	32,116	0	651				651	0.01987
5	0	63					63	1
6	1						0	0

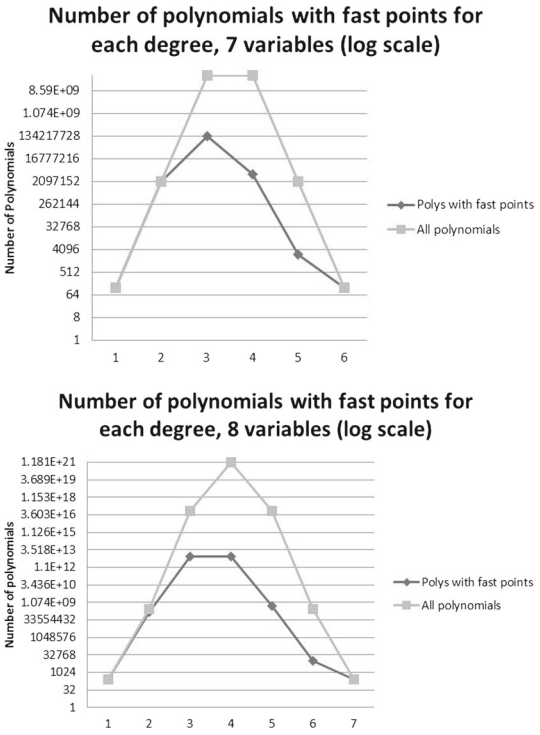

Fig. 1 Number of polynomials with fast points in 7 and 8 variables

Proof Firstly, since V is isomorphic to $\mathbb{F}_p^k$, the quantities $|S_1(k)|$ and $|S_0(k)|$ only depend on k and do not depend on V.

Pick an arbitrary $\mathbf{a} \in V \setminus \{0\}$. Obviously some spanning sets of V will contain $\mathbf{a}$, some will not. We pair each spanning set S that does not contain $\mathbf{a}$ with $S \cup \{\mathbf{a}\}$, which is also a spanning set. Each spanning set appears in at most one such pair. In each pair one of spanning sets is in $S_0(k)$ and other is in $S_1(k)$, so their effect cancels out in $|S_1(k)| - |S_0(k)|$. The spanning sets that do not belong to any pair are precisely the spanning sets S that contain $\mathbf{a}$ but $S \setminus \{\mathbf{a}\}$ is no longer a spanning set for V. Hence $S \setminus \{\mathbf{a}\}$ spans a vector space of dimension $k-1$. There are exactly p^{k-1} subspaces of V of dimension $k-1$ that do not contain $\mathbf{a}$. Each of these spaces has $|S_0(k-1)|$ spanning sets of even cardinality and $|S_1(k-1)|$ spanning sets of odd cardinality. Obviously if S has even cardinality then $S \setminus \{\mathbf{a}\}$ has odd cardinality and the other way around. Summarising, we have

$$|S_1(k)| - |S_0(k)| = p^{k-1}(|S_0(k-1)| - |S_1(k-1)|) = -p^{k-1}(|S_1(k-1)| - |S_0(k-1)|).$$

This recurrence relation together with the initial value $|S_1(1)| - |S_0(1)| = 1$ yields the desired result. □

References

1. Biham, E., Shamir, A.: Differential cryptanalysis of DES-like cryptosystems. J. Cryptol. **4**(1), 3–72 (1991)
2. Dinur, I., Shamir, A.: Cube attacks on tweakable black box polynomials. In: EUROCRYPT, pp. 278–299 (2009)

3. Duan, M., Lai, X.: Higher order differential cryptanalysis framework and its applications. In: International Conference on Information Science and Technology (ICIST), pp. 291–297 (2011)
4. Duan, M., Yang, M., Sun, X., Zhu, B., Lai, X.: Distinguishing properties and applications of higher order derivatives of boolean functions. Inf. Sci. **271**, 224–235 (2014)
5. Knellwolf, S., Meier, W.: High order differential attacks on stream ciphers. Cryptogr. Commun. **4**(3–4), 203–215 (2012)
6. Lai, X.: Higher order derivatives and differential cryptanalysis. In: Blahut, R.E., Costello, D.J. Jr.., Maurer, U., Mittelholzer, T. (eds.) Communications and Cryptography. The Springer International Series in Engineering and Computer Science, vol. 276, pp. 227–233. Springer Verlag (1994)
7. Vielhaber, M.: Breaking ONE.FIVIUM by AIDA an algebraic IV differential attack. Cryptology ePrint Archive, Report 2007/413. http://eprint.iacr.org/ (2007)