DOI 10.1007/s10559-014-9673-x
Cybernetics and Systems Analysis, Vol. 50, No. 6, November, 2014

CYBERNETICS

ALGEBRAICALLY DEGENERATE APPROXIMATIONS OF BOOLEAN FUNCTIONS

A. N. Alekseychuk[a†] **and S. N. Konyushok**[a‡]

UDC 519.7

Abstract. *Properties of k-dimensional approximations of Boolean functions are investigated. One of main results is the theorem on the structure of k-dimensional functions whose degree equals d and whose distance from a given Boolean function of n variables is no longer than* $2^{n-d}(1-\varepsilon)$, $1 \le d \le k \le n$, $\varepsilon \in (0,1)$. *This theorem considerably strengthens the well-known P. Gopalan result and makes it possible to considerably increase the efficiency of his algorithm for constructing all the mentioned k-dimensional Boolean functions.*

Keywords: *correlation cryptanalysis, degenerate Boolean function, k-dimensional function, Walsh–Hadamard transform, finding k-dimensional approximations of Boolean functions.*

INTRODUCTION

A Boolean function $g: \{0,1\}^n \to \{0,1\}$ $(0 \le k \le n)$ is called k-dimensional [1, 2] if there are a function $\varphi: \{0,1\}^k \to \{0,1\}$ and an $n \times k$ matrix A over a field consisting of two elements such that, for any $x \in \{0,1\}^n$, the equality $g(x) = \varphi(xA)$ holds true. The function g is called algebraically degenerate if it is k-dimensional for some $k < n$ and nondegenerate otherwise [3–5].

First results on correlation properties of algebraically degenerate Boolean functions were obtained in the 70s of the last century [3]. At present, interest in the investigation of these functions is conditioned by problems of cryptanalysis and coding theory. We note that [6–8] describe a number of attacks on key stream generators of stream ciphers whose complication functions are algebraically degenerate or are close to such functions.

We denote by $B_{n,k}$ the set of all k-dimensional Boolean functions of n variables. In [5], approximations of Boolean functions by functions from the set $B_{n,n-1}$ are investigated; in particular, the distance between an arbitrary function $f: \{0,1\}^n \to \{0,1\}$ and the set of algebraically degenerate functions of n variables is found, a method for finding functions closest to f in the set $B_{n,n-1}$ is specified, and estimates for their orders are obtained (in [4, 5], the order of degeneracy of a function $g \in B_{n,n-1}$ is understood to be the largest number $n-k$ for which g is a k-dimensional function).

In [1, 2, 9–11], k-dimensional approximations of Boolean functions are investigated for all possible values of k and, in [2], functions over an arbitrary finite field are considered.

In [1], a probabilistic algorithm is proposed for the recognition of the property of k-dimensionality. For any function $f: \{0,1\}^n \to \{0,1\}$ specified with the help of an oracle and numbers $k \in \{0,1,\dots,n-1\}$ and $\varepsilon \in (0,1)$, this algorithm allows one to test the hypothesis H_0: $f \in B_{n,k}$ on the following alternative H_1: f is at the distance that is no longer than $2^n \varepsilon$ from the set of k-dimensional functions after executing $O(n 2^{2k} k \varepsilon^{-1})$ binary operations. A more efficient test of k-dimensionality whose complexity equals $O(n 2^k k^2 \varepsilon^{-1})$ binary operations is proposed in [10].

[a]Institute of Special Communication and Information Protection of the National Technical University of Ukraine "Kyiv Polytechnic Institute," Kyiv, Ukraine, †*alex-crypto@mail.ru*; ‡*3tooth@mail.ru*. Translated from Kibernetika i Sistemnyi Analiz, No. 6, pp. 3–14, November–December, 2014. Original article submitted November 5, 2013.

To construct efficient attacks on symmetric cryptosystems, k-dimensional functions should be found that are rather close to a given function $f: \{0,1\}^n \to \{0,1\}$. Here, of greatest interest is the case when f is specified with the help of an oracle, n is large (for example, $n \geq 64$), and k is small (is fixed or slowly increases with increasing n). The efficiency of solving this problem appreciably depends on the distance between the function f and its sought-for approximations.

Let g be a k-dimensional function of n variables whose distance from the function f is no longer than $2^{n-(k+1)}(1-\varepsilon)$, $\varepsilon \in (0,1)$ (note that the function g is uniquely determined by this condition). In [1], a probabilistic algorithm is proposed that makes it possible to determine g from given f, k, and ε with probability no less than $1-\delta$, $\delta \in (0,1)$, using $O(2^{4k} n^2 \varepsilon^{-2} \log(2^{2k} n \delta^{-1}))$ binary operations. In [11], another algorithm is presented whose binary complexity is equal to $O(2^{2k} k^{-2} n^3 \varepsilon^{-2} \delta^{-1} \log(2^{2k} k^{-1} n \delta^{-1} \varepsilon^{-1}))$.

The problem of definition of all functions $g \in B_{n,k}$ that are at a distance that does not exceed $2^{n-k}(1-\varepsilon)$, $\varepsilon \in (0,1)$, from a given function $f: \{0,1\}^n \to \{0,1\}$ is more difficult. A considerable contribution to its solution was made in [2], which is devoted to the investigation of k-dimensional approximations of functions of n variables over an arbitrary finite field. One of main results of the work is a theorem describing the structure of k-dimensional functions whose degree is no higher than d over a field $\mathbf{GF}(q)$ and whose distance from a given function $f: \mathbf{GF}(q)^n \to \mathbf{GF}(q)$ is no longer than $\delta_q(d)(1-\varepsilon)$, where $\delta_q(d)$ is the minimal distance of the Reed–Muller code $\mathbf{RM}_q(n,d)$, $1 \leq d \leq k$, $\varepsilon \in (0,1)$ (see [2, Par. 4]). The present article proves a theorem essentially strengthening this result for the case of Boolean functions. Several statements about properties of k-dimensional approximations of Boolean functions are also obtained that supplement and refine some results of [5, 9].

DEFINITION OF BASIC CONCEPTS AND SOME AUXILIARY RESULTS

We denote by V_n the set of binary vectors of length n. This set is a vector space of dimension n over the field $F = \mathbf{GF}(2)$ (when $n = 0$, we assume that $V_0 = \{0\}$). The sum of vectors $\alpha = (\alpha_1, \ldots, \alpha_n)$, $x = (x_1, \ldots, x_n) \in V_n$ is found by the formula $\alpha \oplus x = (\alpha_1 \oplus x_1, \ldots, \alpha_n \oplus x_n)$, and the Boolean scalar product is found by the formula $\alpha x = \alpha_1 x_1 \oplus \cdots \oplus \alpha_n x_n$ (here and below, the symbol $\oplus$ denotes the operation of addition of elements of the field F and vectors over this field).

Hereafter, the following notations are used: $\# M$ is the cardinality of a set M; $< M >$ is the subspace generated by a set $M \subseteq V_n$; $F_{n \times k}$ is the set of matrices of size $n \times k$ over the field F; $C(A)$ is the subspace generated by columns of a matrix $A \in F_{n \times k}$ in the vector space V_n.

For any set $M \subseteq V_n$, we denote by $M^{\perp}$ the following subspace dual to M: $M^{\perp} = \{\alpha \in V_n \mid \forall x \in M: \alpha x = 0\}$; for any $a, b \in \mathbf{Z}$, we put $\overline{a,b} = \{i \in \mathbf{Z}: a \leq i \leq b\}$.

We denote by B_n the set of Boolean functions of n variables. The relative distance between functions $f, g \in B_n$ is determined by the formula $d(f,g) = 2^{-n} \# \{x \in V_n: f(x) \neq g(x)\}$, and the relative distance from a function $f \in B_n$ to a set $U \subseteq B_n$ is found by the formula $d(f,U) = \min\limits_{g \in U} d(f,g)$. The number $wt(f) = 2^{-n} \# \{x \in V_n: f(x) = 1\}$ is called the relative weight of a function $f \in B_n$. A Boolean function is called equilibrated if its relative weight is equal to $1/2$.

Denote by $\deg f$ the degree of the Zhegalkin polynomial of a function $f \in B_n$. The proof of the following lemma can be found in [12, Theorem 5.5].

LEMMA 1. Let $f \in B_n \setminus \{0\}$. Then $wt(f) \geq 2^{-\deg f}$.

For any function $f \in B_n$, we put

$$\hat{f}(\alpha) = 2^{-n} \sum_{x \in V_n} (-1)^{f(x) \oplus \alpha x}, \quad \alpha \in V_n, \tag{1}$$

$$D_\alpha f(x) = f(x \oplus \alpha) \oplus f(x), \quad x \in V_n, \tag{2}$$

$$I_f = \{\alpha \in V_n: D_\alpha f \equiv 0\}.$$

The numbers in formula (1) are called Walsh–Hadamard normalized coefficients of the function f, and function (2) is called its directional derivative along a direction $\alpha \in V_n$ [12].