Best Quadratic Approximations of Cubic Boolean Functions

Nicholas Kolokotronis[1,2], Konstantinos Limniotis[1], and Nicholas Kalouptsidis[1]

[1] Department of Informatics and Telecommunications
National and Kapodistrian University of Athens
TYPA Buildings, University Campus, 15784 Athens, Greece
{nkolok, klimn, kalou}@di.uoa.gr

[2] Department of Computer Science and Technology
University of Peloponnese
End of Karaiskaki Street, 22100 Tripolis, Greece
nkolok@uop.gr

Abstract. The problem of computing best low order approximations of Boolean functions is treated in this paper. We focus on the case of best quadratic approximations of a wide class of cubic functions of arbitrary number of variables, and provide formulas for their efficient calculation. Our methodology is developed upon Shannon's expansion formula and properties of best affine approximations of quadratic functions, for which we prove formulas for their direct computation, without use of the Walsh-Hadamard transform. The notion of nonquadricity is introduced, as the minimum distance from all quadratic functions, and cubic functions that achieve the maximum possible nonquadricity are determined, leading to a lower bound for the covering radius of second order Reed-Muller code $\mathfrak{R}(2,n)$ in $\mathfrak{R}(3,n)$.

Key words: Bent functions; boolean functions; covering radius; lower order approximations; nonlinearity; nonquadricity; Reed-Muller codes.

1 Introduction

Boolean functions are used in many different areas and play a prominent role in the security of cryptosystems. Their most important cryptographic applications are in the analysis and design of s-boxes for block ciphers, as well as, filter and combining functions for stream ciphers [33]. In general, resistance of cryptosystems to various cryptanalytic attacks is associated with properties of the Boolean functions used. Apart from their algebraic degree, the *nonlinearity* of Boolean functions is one of the most significant cryptographic properties; it is defined as the minimum distance from all affine functions, and indicates the degree to which attacks based on linear

cryptanalysis [31], and best affine approximations [15], can be prevented. For an even number n of variables, the maximum possible nonlinearity is equal to $2^{n-1} - 2^{n/2-1}$ and this can only be attained by the so-called *bent functions*. Binary bent functions were introduced in [38] and subsequently generalized to the q-ary case in [26]. It is well-known that they correspond to the characteristic function of elementary Hadamard difference sets [14]. Due to their importance, bent functions have received a lot of attention in the past years, and a large number of constructions have been proposed in the literature [3, 4, 11, 14, 16, 19, 20, 38]. On the other hand, the maximum possible nonlinearity attained for odd number n of variables still remains open problem. Apart from the above, many other criteria have also been studied in order to construct cryptographically strong Boolean functions [32, 37, 39]. With the appearance of more recent attacks, such as algebraic [13], and low order approximation attacks [24, 27, 28], Boolean functions need also have the property that they cannot be approximated efficiently by low degree functions. Hence, the *rth order nonlinearity* characteristics of Boolean functions need also be analyzed. This is known to be a difficult task for $r > 1$, whereas even the second order nonlinearity is unknown for all Boolean functions, with the exception of some special cases, or if the number of variables n is small [9]. This problem has also been studied in [34, 35], and an algorithm to determine good rth order approximations (not necessarily best) by repetitive Hamming sphere sampling was given. Proving lower bounds on the rth order nonlinearity of Boolean functions is also considered as a difficult task, even when $r = 2$ [9]. Currently, only few lower bounds have been derived on the rth order nonlinearity [7, 8, 21], but they are quite small. Upper bounds on the rth order nonlinearity have also been derived and are given in [5].

In this paper, we mainly focus on the case of efficiently computing the best quadratic approximations of cubic Boolean functions, leading to the generalizations of many notions and properties that are familiar from the best affine approximation case. More precisely, the nonlinearity has been extended to the *nonquadricity*, which is defined as the minimum distance from all quadratic functions. The cubic functions are classified into classes associated with some integer m; it is shown that m plays a role similar to the rank h of the symplectic matrix corresponding to quadratic functions (roughly speaking, nonquadricity grows with m). Explicit formulas have been proved that compute all best affine (resp. quadratic) approximations of quadratic (resp. cubic) functions, without use of the Walsh-Hadamard transform; this was made possible by proving properties of the best affine approximations, and using Shannon's expansion formula. Cubic functions

that achieve the maximum possible nonquadricity are determined, leading to a lower bound for the covering radius of second order Reed-Muller code $\mathfrak{R}(2,n)$ in $\mathfrak{R}(3,n)$, which is close to the recent upper bound given in [5]. These results hold for an arbitrary number n of variables and lead to constraints on the proper choice of Boolean functions that are stronger than that of *normality* [4, 10, 16]. It is important to note that several constructions of cryptographic primitives based on cubic and quadratic functions have been proposed in the literature (*see* e.g. [18] and [1] respectively) due to their efficient implementation; hence, our results are of cryptographic value when such functions need to be applied.

The paper is organized as follows: Section 2 provides the background and introduces the notation. The properties of best affine approximations of quadratic functions are treated in Section 3, whereas Section 4 studies best quadratic approximations of cubic Boolean functions and determines efficient ways for their computation. Concluding remarks and further research are given in Section 5.

2 Preliminaries

Let $\mathbb{F}_2 = \{0,1\}$ be the finite field of two elements and let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function on n variables mapping elements of the nth dimensional vector space $\mathbb{F}_2^n$ onto $\mathbb{F}_2$ [29]. Boolean functions are commonly expressed in their *algebraic normal form* (ANF)

$$f(x_1, \ldots x_n) = \sum_{j \in \mathbb{F}_2^n} a_j \, x_1^{j_1} \cdots x_n^{j_n} \,, \qquad a_j \in \mathbb{F}_2 \tag{1}$$

where $j = (j_1, \ldots, j_n)$ and the sum is performed modulo 2. The *algebraic degree* of function f is defined as $\deg(f) = \max\{\mathrm{wt}(j) : a_j = 1\}$, where $\mathrm{wt}(j)$ denotes the Hamming weight of vector j. When $\deg(f) = 1, 2$, or 3, then f is said to be an *affine*, *quadratic*, or *cubic* function respectively. Affine functions with zero constant term are called *linear*. In general, the terms of degree $k \leq \deg(f)$ that appear in (1) will be referred to as the *kth degree part* of function f. In the sequel, the set of Boolean functions on n variables is denoted by $\mathbb{B}_n$. The truth table of $f \in \mathbb{B}_n$ is the vector

$$f = \big(f(0,0,\ldots,0), f(1,0,\ldots,0), \ldots, f(1,1,\ldots,1)\big)$$

of length 2^n, which is also denoted by f for simplicity. It is well-known that if $\deg(f) \leq r$, then vector f is a codeword of the rth order binary Reed-Muller code $\mathfrak{R}(r,n)$ [30]. The Hamming weight of the Boolean function

f is equal to the number of the nonzero terms in its truth table, and is said to be *balanced* when it holds $\mathrm{wt}(f) = 2^{n-1}$. Moreover, the *Hamming distance* between two functions $f, g \in \mathbb{B}_n$ is defined as $\mathrm{wt}(f+g)$. A Boolean function $f \in \mathbb{B}_n$ admits the decomposition given below that is used extensively throughout the text.

Definition 1. *Let $j_1, \ldots, j_k$ be integers such that $1 \leq j_1 < \cdots < j_k \leq n$ and $k < n$. Further, let $r = r_1 + 2r_2 + \cdots + 2^{k-1}r_k$ be the binary expansion of the integer $0 \leq r < 2^k$. The expression*

$$f(x_1, \ldots, x_n) = \sum_{r=0}^{2^k-1} \left(\prod_{i=1}^{k} (x_{j_i} + \bar{r}_i) \right) f_r \tag{2}$$

where $\bar{r}$ denotes the complement of r, and each $f_r \in \mathbb{B}_{n-k}$ does not depend on $x_{j_1}, \ldots, x_{j_k}$, is called the kth order Shannon's expansion formula *of f with respect to the variables $x_{j_1}, \ldots, x_{j_k}$.*

It is clear by Definition 1 that for any choice of $x_{j_1}, \ldots, x_{j_k}$, the functions $f_0, \ldots, f_{2^k-1}$, called *sub-functions*, are uniquely defined, as f_r is obtained from f by setting $x_{j_i} = r_i$, $1 \leq i \leq k$. Let us write $\mathcal{J} = \{j_1, \ldots, j_k\}$; then, the kth order Shannon's expansion formula, given by (2), will be denoted by $f = f_0 \parallel_{\mathcal{J}} \cdots \parallel_{\mathcal{J}} f_{2^k-1}$ (when $\mathcal{J} = \{j\}$ we simply write $f = f_0 \parallel_j f_1$). If $\mathcal{J} = \{n-k+1, \ldots, n\}$, the truth table of $f(x_1, \ldots, x_n)$ is constructed by *concatenating* the truth tables of the sub-functions $f_r(x_1, \ldots, x_{n-k})$ [35]. This case is denoted by $f = f_0 \parallel \cdots \parallel f_{2^k-1}$ and will be referred to as the *kth order Shannon's expansion formula* of f [25].

Remark 1. In the case of $f = f_0 \parallel_j f_1$, the Shannon's expansion formula coincides with the $\mid u \mid u+v \mid$ construction in coding theory [30, p. 76]. Indeed, we can always write

$$f = f_0 + x_j f_1' = (x_j + 1) f_0 + x_j (f_0 + f_1') = f_0 \parallel_j (f_0 + f_1')$$

where $\deg(f_0) \leq \deg(f)$, $\deg(f_1') \leq \deg(f) - 1$, and f_0, f_1 do not depend on the variable x_j. Thus, if we have $f \in \mathfrak{R}(r, n)$ then $f_0 \in \mathfrak{R}(r, n-1)$ and $f_1' \in \mathfrak{R}(r-1, n-1)$. □

The Walsh or Hadamard transform of the Boolean function $f \in \mathbb{B}_n$ at $a \in \mathbb{F}_2^n$, denoted by $\widehat{\chi}_f(a)$, is the real-valued function given by [6]

$$\widehat{\chi}_f(a) = \sum_{x \in \mathbb{F}_2^n} \chi_f(x) (-1)^{\langle a, x \rangle} = 2^n - 2\,\mathrm{wt}(f + \phi_a) \tag{3}$$

where $\chi_f(x) = (-1)^{f(x)}$ is said to be the *sign function* of f and ϕ_a is the linear function $\phi_a(x) = \langle a, x\rangle = a_1x_1 + \cdots + a_nx_n$. It is clear from (3) that the Walsh transform of f corresponds to the Fourier transform of the function χ_f. Furthermore, (3) implies that f is balanced if and only if $\widehat{\chi}_f(0) = 0$. The *Walsh support* of the Boolean function f is defined as $S_f = \left\{a \in \mathbb{F}_2^n : \widehat{\chi}_f(a) \neq 0\right\}$. The minimum distance between f and all affine functions is referred to as the *nonlinearity* of f and is denoted by $\mathcal{NL}_f$; it is determined by the Walsh transform spectra as follows [32, 36]

$$\mathcal{NL}_f = \min_{g \in \mathfrak{R}(1,n)} \left\{\mathrm{wt}(f+g)\right\} = 2^{n-1} - \frac{1}{2} \max_{a \in \mathbb{F}_2^n} |\widehat{\chi}_f(a)| \,. \tag{4}$$

Any affine function g such that $\mathrm{wt}(f+g) = \mathcal{NL}_f$ is called a *best affine approximation* of f and is denoted by λ_f, whereas the set comprising of best affine approximations of f is denoted by $\mathcal{A}_f \subseteq \mathfrak{R}(1,n)$. The definition of the nonlinearity leads directly to the following well-known result.

Lemma 1. *Let $f \in \mathbb{B}_n$ and $h \in \mathfrak{R}(1,n)$. Then, $g + h \in \mathcal{A}_{f+h}$ if and only if $g \in \mathcal{A}_f$. Further, $|\mathcal{A}_{f+h}| = |\mathcal{A}_f|$, i.e. both sets have the same cardinality.*

An equivalent statement of Lemma 1, which is subsequently used, is that $\lambda_{f+h} = \lambda_f + h$ for any linear function h, and a proper choice of λ_f, λ_{f+h}. Similarly, the minimum distance between f and all quadratic functions is called the *second-order nonlinearity* or *nonquadricity* of f and is denoted by $\mathcal{NQ}_f \triangleq \mathcal{NL}_f^2$; it is given by

$$\mathcal{NQ}_f = \min_{g \in \mathfrak{R}(2,n)} \left\{\mathrm{wt}(f+g)\right\} . \tag{5}$$

From (4), (5) we clearly obtain that $\mathcal{NQ}_f \leq \mathcal{NL}_f$. Likewise, any quadratic function g with the property $\mathrm{wt}(f+g) = \mathcal{NQ}_f$ is said to be a *best quadratic approximation* of f and is denoted by ξ_f, whereas the set comprising of best quadratic approximations of f is denoted by $\mathcal{Q}_f \subseteq \mathfrak{R}(2,n)$. The above concepts are easily generalized to include the *rth order nonlinearity* $\mathcal{NL}_f^r$ and *approximation* of the Boolean function f respectively [24].

A well-known relationship exists between the nonlinearity (resp. nonquadricity) of Boolean functions and the covering radius of the first (resp. second) order Reed-Muller code. More precisely, the *covering radius* of the r-th order Reed-Muller code $\mathfrak{R}(r,n)$ equals the smallest integer $\rho = \rho(r,n)$ such that any binary vector of length 2^n lies within Hamming distance ρ from some codeword of $\mathfrak{R}(r,n)$ [12, 30]; it is given by [2, 27, 28]

$$\rho(r,n) = \max_{f \in \mathbb{B}_n} \min_{g \in \mathfrak{R}(r,n)} \left\{\mathrm{wt}(f+g)\right\} . \tag{6}$$

If we confine the Boolean function $f \in \mathbb{B}_n$ into $\mathfrak{R}(s,n)$ in (6), with $s \geq r$, we denote the result by $\rho_s(r,n)$; obviously, it holds $\rho_s(r,n) \leq \rho(r,n)$. By comparing (6) with (4) (resp. (5)) we conclude that $\rho(1,n)$ (resp. $\rho(2,n)$) corresponds to the maximum nonlinearity (resp. nonquadricity) that any Boolean function in $\mathbb{B}_n$ can achieve. In the case of even n, it is well-known that $\rho(1,n) = 2^{n-1} - 2^{n/2-1}$, whereas for odd n we have [17]

$$2^{n-1} - 2^{\frac{n-1}{2}} \leq \rho(1,n) < 2^{n-1} - 2^{\frac{n}{2}-1} .$$

A recent upper bound on the maximum nonquadricity $\rho(2,n)$ of Boolean functions has been proved in [5]. Due to the existence of efficient attacks exploiting low order, i.e. not necessarily affine, approximations of Boolean functions (*see* e.g. [13, 24, 34, 35]), it is very important that they cannot be approximated to a large extent.

3 Properties of Best Affine Approximations

In this section we review some essential properties of quadratic functions and prove results that are subsequently used to derive the best quadratic approximation of Boolean functions having higher algebraic degree. Let $f \in \mathbb{B}_n$ be a quadratic function and let $x = (x_1, \ldots, x_n)$; then, f can be written as $f = xQx^T + Lx^T + \epsilon$ for some upper triangular binary matrix Q, binary vector L, and a constant $\epsilon \in \mathbb{F}_2$, where xQx^T is the quadratic part of f. It is well-known that (*see* e.g. [30, pp. 434–442]) the rank of the symplectic matrix $B = Q + Q^T$ equals $2h$, for some $1 \leq h \leq \lfloor n/2 \rfloor$. By Dickson's theorem there exists a nonsingular matrix $R = (r_{i,j})_{i,j=1}^n$ such that the only nonzero elements of $\tilde{B} = R^{-1}B(R^{-1})^T$ lie in its subdiagonal and superdiagonal (more precisely $\tilde{b}_{2i-1,2i} = \tilde{b}_{2i,2i-1} = 1$, $1 \leq i \leq h$), and under the transformation of variables $g = xR$, the function f becomes

$$f = g_0 + \sum_{i=1}^{h} g_{2i-1}g_{2i} , \qquad \deg(g_0) \leq 1 \text{ and } \deg(g_j) = 1 \qquad (7)$$

with $g_0 = \tilde{g} + \tilde{L}g^T + \epsilon$ for a linear function $\tilde{g}$ derived from the quadratic part of f (its properties are studied in Proposition 2) and $\tilde{L} = L(R^{-1})^T$, whereas $\{g_1, \ldots, g_{2h}\}$ are linearly independent linear functions (actually we have $g_j = \sum_{i=1}^n r_{i,j}x_i$). Since h only depends on the quadratic part of the Boolean function f, it is denoted by h_f; clearly $h_f = 0$ if $f \in \mathfrak{R}(1,n)$. The Walsh spectra of f are fully determined by h_f, as seen below.

Theorem 1 ([30]). *Let $\mathcal{B}_f = \{f + g : g \in \mathfrak{R}(1,n)\}$ for a fixed quadratic Boolean function $f \in \mathfrak{R}(2,n)$. Then*

$$\mathrm{wt}(f+g) = \begin{cases} 2^{n-1} - 2^{n-h_f-1}, & \text{occurs } 2^{2h_f} \text{ times;} \\ 2^{n-1}, & \text{occurs } 2^{n+1} - 2^{2h_f+1} \text{ times;} \\ 2^{n-1} + 2^{n-h_f-1}, & \text{occurs } 2^{2h_f} \text{ times.} \end{cases}$$

From the coding theory point of view, Theorem 1 determines the weight distribution of a coset $\mathcal{B}_f$ of $\mathfrak{R}(1,n)$ in $\mathfrak{R}(2,n)$. Then, according to (4), the nonlinearity of any quadratic function $f \in \mathbb{B}_n$ equals $2^{n-1} - 2^{n-h_f-1}$. For all even integers m, we subsequently introduce the linear mappings $\zeta : \mathbb{F}_2^m \times \mathbb{F}_2^m \to \mathbb{F}_2$ and $\psi : \mathbb{F}_2^m \to \mathbb{F}_2$ as

$$\zeta(x,y) = \sum_{i=1}^{m/2} \begin{vmatrix} x_{2i-1} & x_{2i} \\ y_{2i-1} & y_{2i} \end{vmatrix} = \sum_{i=1}^{m/2} \left(x_{2i-1}y_{2i} + x_{2i}y_{2i-1}\right). \tag{8}$$

and $\psi(x) = \sum_{i=1}^{m/2} x_{2i-1}x_{2i}$ respectively (actually m is equal to the length of their arguments). The following statement, allows to directly compute all best affine approximations of a quadratic function.

Theorem 2. *Let the Boolean function $f \in \mathfrak{R}(2,n)$ be given by (7). Then, for $b = (b_1, \ldots, b_{2h})$ we have that*

$$\mathcal{A}_f = \left\{\lambda_f^b \in \mathfrak{R}(1,n) : \lambda_f^b = g_0 + \sum_{i=1}^{2h} b_i g_i + \psi(b),\ b \in \mathbb{F}_2^{2h}\right\}.$$

Proof. First note that the affine Boolean functions λ_f^b are pairwise distinct since from hypothesis we have that $\{g_1, \ldots, g_{2h}\}$ are linearly independent; thus $|\mathcal{A}_f| = 2^{2h}$. Furthermore, for all $b \in \mathbb{F}_2^{2h}$, the distance of λ_f^b from f is equal to the weight of

$$\begin{aligned} f + \lambda_f^b &= \sum_{i=1}^{h} \left(g_{2i-1}g_{2i} + b_{2i-1}g_{2i-1} + b_{2i}g_{2i} + b_{2i-1}b_{2i}\right) \\ &= \sum_{i=1}^{h} \left(g_{2i-1} + b_{2i}\right)\left(g_{2i} + b_{2i-1}\right). \end{aligned} \tag{9}$$

Since $\{g_1 + b_2, \ldots, g_{2h} + b_{2h-1}\}$ are also linearly independent, we get that for any choice of $b \in \mathbb{F}_2^{2h}$ it holds $\mathrm{wt}(f + \lambda_f^b) = 2^{n-1} - 2^{n-1-h}$ [30], which by Theorem 1 is the minimum distance between f and all affine functions. The fact that the number of best affine approximations of f is 2^{2h} (equal to the number of different λ_f^b constructed here) concludes our proof. □

Example 1. Let $f \in \mathbb{B}_5$ be the quadratic function given by $f(x_1, \ldots, x_5) = x_1x_2 + x_1x_5 + x_2x_3 + x_3x_5 + x_2 + x_4$. From the above analysis, it can be easily found that the Boolean function f is written in the following form $f(x_1, \ldots, x_5) = (x_1 + x_3)(x_2 + x_5) + x_2 + x_4$ and according to Theorem 2 its best affine approximations are

$$\lambda_f^0 = x_2 + x_4\,, \qquad \lambda_f^1 = x_1 + x_2 + x_3 + x_4\,,$$
$$\lambda_f^2 = x_4 + x_5\,, \qquad \lambda_f^3 = x_1 + x_3 + x_4 + x_5 + 1\,.$$

Note that one of the solutions is the linear part of f; as shown next, this can be directly found by examining the weight of its quadratic part. □

Corollary 1. *Let the Boolean function $f \in \mathfrak{R}(2, n)$ be given by (7), and let $\mathcal{P}_{2h}$ be the set of permutations of $\{1, \ldots, 2h\}$. Then, for any $\pi \in \mathcal{P}_{2h}$, the best affine approximations of $f_\pi = g_0 + \sum_{i=1}^{h} g_{\pi_{2i-1}}\, g_{\pi_{2i}}$ are given by*

$$\lambda_{f_\pi}^b = \lambda_f^{b_\sigma} + \psi(b) + \psi(b_\sigma)\,, \qquad \forall\ b \in \mathbb{F}_2^{2h}$$

where $\sigma = \pi^{-1}$ and $b_\sigma = (b_{\sigma_1}, \ldots, b_{\sigma_{2h}})$.

Proof. This is a direct result of Theorem 2, since for all $b \in \mathbb{F}_2^{2h}$ we have $\lambda_{f_\pi}^b + \psi(b) = g_0 + \sum_{i=1}^{2h} b_i\, g_{\pi_i} = g_0 + \sum_{i=1}^{2h} b_{\sigma_i}\, g_i = \lambda_f^{b_\sigma} + \psi(b_\sigma)$. □

Proposition 1. *Let $f \in \mathfrak{R}(2, n)$ be given by $f = q + l$, where q, l are its quadratic and linear part respectively. Then, q is not balanced if and only if $l + \epsilon$ is a best affine approximation of f for some $\epsilon \in \mathbb{F}_2$.*

Proof. From Theorem 1, we have that $\widehat{\chi}_q(a) \in \left\{0, \pm 2^{n-h_f}\right\}$ for all $a \in \mathbb{F}_2^n$, where $\widehat{\chi}_q(0) \neq 0$ if and only if q is not balanced, due to (3). Hence, from $2^n - 2\,\mathrm{wt}(f) = \widehat{\chi}_q(0) = (-1)^\epsilon\, 2^{n-h_f}$ we get $\mathrm{wt}(f) = 2^{n-1} - (-1)^\epsilon\, 2^{n-1-h_f}$ and $\lambda_q = \epsilon$ is a best affine approximation of q. Subsequent application of Lemma 1 yields that $\lambda_f = l + \lambda_q$ is a best affine approximation of f. □

Proposition 2. *With the above notation, let R be the nonsingular matrix such that $f = xQx^T + Lx^T + \epsilon$ is given by (7) under the transformation of variables $g = xR$. Moreover, let vector r_i contain the first $2h$ elements of the ith row of matrix R, that is $r_i = (r_{i,1}, \ldots, r_{i,2h})$. Then, we have*

1. *The linear function $\tilde{g}$, resulting from the quadratic part of f, is given by $\tilde{g} = \sum_{i=1}^{n} \psi(r_i)\, x_i$;*
2. *The upper triangular matrix $Q = (q_{i,j})_{i,j=1}^{n}$ is related with R by means of $q_{i,j} = \zeta(r_i, r_j)$;*

3. *Moreover, the quadratic part of f is balanced if and only if $\tilde{g} \neq 0$ and linearly independent of $\{g_1, \ldots, g_{2h}\}$.*

Proof. From hypothesis, the quadratic part xQx^T of function f becomes $\tilde{g} + \sum_{j=1}^{h} g_{2j-1}\, g_{2j}$ under the transformation of variables $g = xR$, where the linear functions g_j are given by $g_j = \sum_{i=1}^{n} r_{i,j}\, x_i$. By substituting g_j at the right hand-side of the following equality we obtain

$$\begin{aligned} xQx^T + \tilde{g} &= \sum_{j=1}^{h} g_{2j-1}\, g_{2j} = \sum_{j=1}^{h} \left(\sum_{i=1}^{n} r_{i,2j-1}\, x_i \right) \left(\sum_{k=1}^{n} r_{k,2j}\, x_k \right) \\ &= \sum_{j=1}^{h} \left(\sum_{i=1}^{n-1} \sum_{k=i+1}^{n} \begin{vmatrix} r_{i,2j-1} & r_{i,2j} \\ r_{k,2j-1} & r_{k,2j} \end{vmatrix} x_i x_k + \sum_{i=1}^{n} r_{i,2j-1} r_{i,2j}\, x_i \right) \\ &= \sum_{i=1}^{n-1} \sum_{k=i+1}^{n} \zeta(r_i, r_k)\, x_i x_k + \sum_{i=1}^{n} \psi(r_i)\, x_i \end{aligned} \tag{10}$$

which establishes the first two properties. In fact, property 1 states that x_i is present in the linear function $\tilde{g}$ if and only if it is a common variable of g_{2j-1} and g_{2j} for odd number of $1 \leq j \leq h$. To prove the last property, note that if $\tilde{g} \neq 0$ and $\tilde{g} = \sum_{j=1}^{2h} a_j\, g_j$, i.e. $a_j = 0$ for all $j > 2h$, then $\tilde{g}$ linearly depends on $\{g_1, \ldots, g_{2h}\}$, and the quadratic part of function f is written as $\sum_{j=1}^{h} \big(g_{2j-1} + a_{2j}\big)\big(g_{2j} + a_{2j-1}\big) + \psi(a)$, which is not balanced (*see* the proof of Theorem 2). On the other hand, if we have that $a_j = 1$ for some $j > 2h$, then xQx^T is balanced since g_j does not linearly depend on $\{g_1, \ldots, g_{2h}\}$ due to the invertibility of R [30, p. 442]. □

Example 2. Let $f \in \mathbb{B}_5$ be the quadratic function given by $f(x_1, \ldots, x_5) = x_1x_3 + x_1x_5 + x_3x_5 + x_2 + x_4$. From the above analysis, it is found that f becomes $f(x_1, \ldots, x_5) = (x_1 + x_5)(x_3 + x_5) + x_2 + x_4 + x_5$, and according to Theorem 2 its best affine approximations are

$$\begin{aligned} \lambda_f^0 &= x_2 + x_4 + x_5\,, & \lambda_f^1 &= x_1 + x_2 + x_4\,, \\ \lambda_f^2 &= x_2 + x_3 + x_4\,, & \lambda_f^3 &= x_1 + x_2 + x_3 + x_4 + x_5 + 1\,. \end{aligned}$$

Note that the new term $\tilde{g} = x_5$ added to the linear part of f is indeed the common variable contained in $g_1 = x_1 + x_5$ and $g_2 = x_3 + x_5$; hence, according to Proposition 2 the quadratic part of f is balanced. □

Before we prove the following statement, we first introduce a commonly used partial ordering of elements of the vector space $\mathbb{F}_2^n$. For all $a, b \in \mathbb{F}_2^n$ we write $a \preceq b$ if and only if $a_i \leq b_i$ for all $1 \leq i \leq n$. The relation $a \succeq b$

is similarly defined. Next, we prove that the best affine approximations of a quadratic function, given as the sum of linearly independent quadratic functions, can be computed in terms of their best affine approximations.

Theorem 3. *Let the Boolean functions $f_1, f_2 \in \mathfrak{R}(2,n)$ be given by (7). Then for some properly chosen fixed binary vector b', of length $2h_{f_1}+2h_{f_2}$ and weight $2h_{f_1+f_2}$, we have*

$$\lambda^b_{f_1+f_2} = \lambda^{c_1}_{f_1} + \lambda^{c_2}_{f_2} + \epsilon(b)\,, \quad \forall\ b \preceq b' \tag{11}$$

where $c_i \triangleq c_i(b) \in \mathbb{F}_2^{2h_{f_i}}$ and $\epsilon(b) \in \mathbb{F}_2$ depend on vector b.

Proof. Let $f = f_1 + f_2$; we write h and h_i instead of h_f and h_{f_i} to simplify notation. From hypothesis we get $f = \sum_{i=1}^{2} g_{i,0} + \sum_{i=1}^{2}\sum_{j=1}^{h_i} g_{i,2j-1}g_{i,2j}$, where for $i = 1, 2$ the Boolean functions in the set $\mathcal{G}_i = \{g_{i,1}, \ldots, g_{i,2h_i}\}$ are linearly independent. By considering functions $g_{i,j}$ as elements of the vector space $\mathbb{F}_2^n$ ($g_{i,j}(x) = \phi_a(x)$ for some $a \in \mathbb{F}_2^n$), we define the mapping

$$\mathbb{F}_2^{2h_1+2h_2} \ni (b_1, b_2) = b \ \stackrel{L}{\longmapsto}\ L(b) = \sum_{i=1}^{2}\sum_{j=1}^{2h_i} b_{i,j}\, g_{i,j} \in \mathbb{F}_2^n\,. \tag{12}$$

Let d be the nullity of L, that is let $d = \dim\ker(L)$, where $\ker(L)$ is the kernel or null space of the mapping L. From the above, we obviously get $\max\{0, h_1 + h_2 - \frac{n}{2}\} \le d \le \min\{|\mathcal{G}_1|, |\mathcal{G}_2|\} = 2\min\{h_1, h_2\} \le h_1 + h_2$. In the sequel, we assume without loss of generality that $h_1 + h_2 \le n/2$.

Let us have $d = 0$; this implies that the Boolean functions in $\mathcal{G}_1 \cup \mathcal{G}_2$ are linearly independent. Hence, we have $h = h_1 + h_2$, and by Theorem 2 the best affine approximations of f are given by

$$\lambda^b_f = \sum_{i=1}^{2} g_{i,0} + \sum_{i=1}^{2}\sum_{j=1}^{2h_i} b_{i,j}\, g_{i,j} + \psi(b) = \lambda^{b_1}_{f_1} + \lambda^{b_2}_{f_2} \tag{13}$$

for all $b = (b_1, b_2) \in \mathbb{F}_2^{2h}$, since $\psi(b) = \psi(b_1) + \psi(b_2)$.

Next, suppose $d = 1$. Then, there exists $a = (a_1, a_2) \in \mathbb{F}_2^{2h_1+2h_2} \setminus \{0\}$ such that $L(a) = 0$; obviously, both a_1, a_2 are nonzero since any equation of the form $L(a_1, 0) = 0$ or $L(0, a_2) = 0$ necessarily leads to $a_1 = 0$ and $a_2 = 0$; otherwise, the functions in $\mathcal{G}_1$ or $\mathcal{G}_2$ would be linearly dependent, contradicting hypothesis. Let $a_{k,2l-e} = 1$ for some integers $1 \le k \le 2$ and $1 \le l \le h_k$, with $e \in \mathbb{F}_2$. Hence, all functions in $\big(\mathcal{G}_1 \cup \mathcal{G}_2\big) \setminus \{g_{k,2l-e}\}$ are

linearly independent. By substituting $g_{k,2l-e}$ in f from $L(a)=0$ we have

$$\begin{aligned} f &= \sum_{i=1}^{2} g_{i,0} + \sum_{\substack{i=1 \\ (i,j)\neq(k,l)}}^{2} \sum_{j=1}^{h_i} g_{i,2j-1}\, g_{i,2j} + \left(\sum_{\substack{i=1 \\ (i,j)\neq(k,2l-e)}}^{2} \sum_{j=1}^{2h_i} a_{i,j}\, g_{i,j} \right) g_{k,2l-\bar{e}} \\ &= \sum_{i=1}^{2} g_{i,0} + \sum_{\substack{i=1 \\ (i,j)\neq(k,l)}}^{2} \sum_{j=1}^{h_i} g'_{i,2j-1}\, g'_{i,2j} + \psi(a)\, g_{k,2l-\bar{e}} \end{aligned} \tag{14}$$

where $\bar{e}$ is the complement of e and $g'_{i,2j-s} = g_{i,2j-s} + a_{i,2j-\bar{s}}\, g_{k,2l-\bar{e}}$, for $s \in \mathbb{F}_2$. According to the above the functions $g'_{i,j}$ in the quadratic part of f are also linearly independent, hence giving $h = h_1 + h_2 - 1$. Let vector $b' = \left(b'_1, b'_2\right) \in \mathbb{F}_2^{2h_1+2h_2}$ be defined as $b'_{k,2l-1} = b'_{k,2l} = 0$ and $b'_{i,j} = 1$ in all other cases. From (14) and Theorem 2, simple calculations yield that the best affine approximations of f, for all 2^{2h} vectors $b \preceq b'$, are

$$\begin{aligned} \lambda_f^b &= \sum_{i=1}^{2} g_{i,0} + \sum_{\substack{i=1 \\ (i,j)\neq(k,l)}}^{2} \sum_{j=1}^{h_i} \left(b_{i,2j-1}\, g'_{i,2j-1} + b_{i,2j}\, g'_{i,2j} \right) + \psi(a)\, g_{k,2l-\bar{e}} + \psi(b) \\ &= \sum_{i=1}^{2} g_{i,0} + \sum_{i=1}^{2} \sum_{j=1}^{h_i} \left(b_{i,2j-1}\, g_{i,2j-1} + b_{i,2j}\, g_{i,2j} + \begin{vmatrix} a_{i,2j-1} & a_{i,2j} \\ b_{i,2j-1} & b_{i,2j} \end{vmatrix} g_{k,2l-\bar{e}} \right) \\ &\quad + \psi(a)\, g_{k,2l-\bar{e}} + \psi(b_1) + \psi(b_2) \end{aligned}$$

by substituting $g'_{i,2j-1}, g'_{i,2j}$ and using the fact that $b_{k,2l-1} = b_{k,2l} = 0$. It is now readily established that

$$\lambda_f^b = \lambda_{f_1}^{b_1} + \lambda_{f_2}^{b_2} + \left(\psi(a) + \zeta(a,b)\right) g_{k,2l-\bar{e}} = \lambda_{f_1}^{c_1} + \lambda_{f_2}^{c_2} \tag{15}$$

where only one of b_1, b_2 is modified; more precisely c_k is equal to b_k with the exception of $c_{k,2l-\bar{e}} = \psi(a) + \zeta(a,b)$. Note that since $c_{k,2l-e} = 0$ (due to $b_{k,2l-e} = 0$) we get that $\psi(c_k) = \psi(b_k)$, which allows to incorporate the term $c_{k,2l-\bar{e}}\, g_{k,2l-\bar{e}}$ into $\lambda_{f_k}^{b_k}$ and obtain $\lambda_{f_k}^{c_k}$ in (15).

In the case $d \geq 2$, there exist d vectors $a^i = (a_1^i, a_2^i) \in \mathbb{F}_2^{2h_1+2h_2} \setminus \{0\}$ such that $L(a^i) = 0$, for $1 \leq i \leq d$, and $\{a^1, \ldots, a^d\}$ is a basis of $\ker(L)$. Let A be the $d \times 2(h_1 + h_2)$ matrix whose ith row is vector a^i; it is clear that we may write A in the following block form

$$A = (A_1, A_2) = (A_{1,1}, \ldots, A_{1,h_1}, A_{2,1}, \ldots, A_{2,h_2})$$

where A_k is the $d \times 2h_k$ matrix whose ith row is vector a_k^i, for $k = 1, 2$, and each $A_{k,l}$ has size $d \times 2$. From (12) and the above discussion, we see that the d functions $g_{k_1,2l_1-e_1}, \ldots, g_{k_d,2l_d-e_d}$ we will exclude from $\mathcal{G}_1 \cup \mathcal{G}_2$ are associated with the nonzero blocks $A_{k_1,l_1}, \ldots, A_{k_d,l_d}$ of matrix A. As shown next in Remark 2, these d functions can be chosen such that they correspond to d linearly independent columns of matrix A, from distinct blocks A_{k_j,l_j}, and we can assume that $A' = (a^i_{k_j,2l_j-e_j})_{i,j=1}^d$ is the identity matrix without any loss of generality. Let us introduce the following sets of distinct elements $\mathcal{E}_d = \{(k_1, l_1), \ldots, (k_d, l_d)\}$, and

$$\mathcal{E}_d^0 = \{(k_j, 2l_j - e_j) : 1 \le j \le d\}\,, \quad \mathcal{E}_d^1 = \{(k_j, 2l_j - \bar{e}_j) : 1 \le j \le d\}\,.$$

From the preceding analysis, we conclude that all the functions in the set $\big(\mathcal{G}_1 \cup \mathcal{G}_2\big) \setminus \big(\bigcup_{(i,j)\,\in\,\mathcal{E}_d^0}\{g_{i,j}\}\big)$ are linearly independent. Working similar to the case $d = 1$, substitution of $g_{k_i,2l_i-e_i}$ in f using the equation $L(a^i) = 0$ will lead to the following expression

$$f = \sum_{i=1}^{2} g_{i,0} + \sum_{\substack{i=1\\(i,j)\,\notin\,\mathcal{E}_d}}^{2}\sum_{j=1}^{h_i} g'_{i,2j-1}\, g'_{i,2j} + \sum_{i=1}^{d-1}\sum_{j=i+1}^{d} \zeta(a^i, a^j)\, y_i\, y_j + \sum_{i=1}^{d} \psi(a^i)\, y_i \tag{16}$$

where $g'_{i,2j-s} = g_{i,2j-s} + \sum_{t=1}^{d} a^t_{i,2j-\bar{s}}\, y_t$, for $s \in \mathbb{F}_2$, and $y_i \triangleq g_{k_i,2l_i-\bar{e}_i}$. In this case, new quadratic terms enter f and therefore h is not necessarily equal to $h_1 + h_2 - d$. Since $\{y_1, \ldots, y_d\}$ are linearly independent, then by Proposition 2 and Dickson's theorem there exists a linear transformation $y' = yP$ (the matrix P depends on the choice of the kernel basis, whereas we have that $y'_j = \sum_{i=1}^{d} p_{i,j}\, y_i$) such that f becomes

$$f = \sum_{i=1}^{2} g_{i,0} + \sum_{\substack{i=1\\(i,j)\,\notin\,\mathcal{E}_d}}^{2}\sum_{j=1}^{h_i} g'_{i,2j-1}\, g'_{i,2j} + \sum_{i=1}^{m} y'_{2i-1}\, y'_{2i} + \sum_{i=1}^{d} \big(\psi(p_i) + \psi(a^i)\big) y_i$$

where $0 \le m \le \lfloor d/2 \rfloor$ and the vector p_i contains the first $2m$ elements of the ith row of matrix P, that is $p_i = (p_{i,1}, \ldots, p_{i,2m})$. Note that $m = 0$ if all $\zeta(a^i, a^j)$ are zero, whereas by Proposition 2 we get $\zeta(p_i, p_j) = \zeta(a^i, a^j)$. Since $\{y'_1, \ldots, y'_{2m}\}$ are also linearly independent from the $2(h_1 + h_2 - d)$ functions $g'_{i,j}$ present in f, we have $h = h_1 + h_2 - (d - m)$.

Let $\widehat{\mathcal{E}}_d^1$ and $\widetilde{\mathcal{E}}_d^1$ contain the first $2m$ and last $d - 2m$ elements of the set $\mathcal{E}_d^1$ respectively. In order to compute, using Theorem 2, the best affine approximations of f we define vector $b' = \big(b'_1, b'_2\big) \in \mathbb{F}_2^{2h_1+2h_2}$ as follows:

$b'_{i,j} = 0$ if $(i,j) \in \mathcal{E}_d^0 \cup \widetilde{\mathcal{E}}_d^1$ and $b'_{i,j} = 1$ otherwise. Then for all 2^{2h} vectors $b \preceq b'$, we find by using arguments similar to the case $d = 1$ that the best affine approximations of f are given by

$$\lambda_f^b = \lambda_{f_1}^{b_1} + \lambda_{f_2}^{b_2} + \sum_{i=1}^{d} \left(\psi(p_i) + \psi(a^i) + \zeta(a^i, b) + \sum_{j=1}^{2m} p_{i,j}\, b_{k_j, 2l_j - \bar{e}_j} \right) \times g_{k_i, 2l_i - \bar{e}_i} + \psi(b_{k_1, 2l_1 - \bar{e}_1}, \ldots, b_{k_{2m}, 2l_{2m} - \bar{e}_{2m}})$$

where both terms $\zeta(a^i, b)$ and $\sum_{j=1}^{2m} p_{i,j}\, b_{k_j, 2l_j - \bar{e}_j}$ result from substituting $g'_{i,2j-1}, g'_{i,2j}$ and y'_{2j-1}, y'_{2j} with their respective expressions. Let $b(\mathcal{E}_d^1)$ and $u(b)$ be the $1 \times d$ vectors whose ith element is given by $b_i(\mathcal{E}_d^1) = b_{k_i, 2l_i - \bar{e}_i}$, i.e. it contains all $b_{i,j}$ with $(i,j) \in \mathcal{E}_d^1$, and $u_i(b) = \psi(p_i) + \psi(a^i) + \zeta(a^i, b)$; vectors $b(\widehat{\mathcal{E}}_d^1)$ and $b(\widetilde{\mathcal{E}}_d^1)$ are similarly defined. From the above formula we see that only elements of $b(\mathcal{E}_d^1)$ are modified, whilst the update function (due to the fact that the last $d - 2m$ elements of $b(\mathcal{E}_d^1)$ are zero) is

$$c(\mathcal{E}_d^1) = b(\mathcal{E}_d^1) \cdot \left(I + P^T \right) + u(b) \tag{17}$$

where I is the identity matrix of order d and vector $c = (c_1, c_2)$ is equal to b for all $(i,j) \notin \mathcal{E}_d^1$. Thus $\lambda_f^b = \lambda_{f_1}^{c_1} + \lambda_{f_2}^{c_2} + \epsilon(b)$, where $\epsilon(b) = \psi(b(\widehat{\mathcal{E}}_d^1))$, which concludes our proof. □

Remark 2. In the proof of Theorem 3, we can easily show (as done in the case $d = 1$) that both parts a_1^i, a_2^i of all vectors a^i in the basis $\{a^1, \ldots, a^d\}$ of $\ker(L)$ are nonzero, due to the linear independence of the functions in $\mathcal{G}_1$ and $\mathcal{G}_2$. In fact, this property holds for all nonzero linear combinations $\tilde{a}^r = (\tilde{a}_1^r, \tilde{a}_2^r)$ of the kernel basis elements (for the same reason), where

$$\tilde{a}^r = \sum_{i=1}^{d} r_i\, a^i \;\Leftrightarrow\; \left(\tilde{a}_1^r, \tilde{a}_2^r\right) = \left(\sum_{i=1}^{d} r_i\, a_1^i\,,\; \sum_{i=1}^{d} r_i\, a_2^i \right), \qquad r \in \mathbb{F}_2^d\,.$$

The above implies that vectors $\{a_k^1, \ldots, a_k^d\}$ are also linearly independent, for $k = 1, 2$; hence, there exists *at least one set* of d linearly independent columns in the matrix $A_k = (A_{k,1}, \ldots, A_{k,h_k})$, for $k = 1, 2$. We can always choose d linearly independent columns from $A = (A_1, A_2)$ such that they belong to distinct nonzero blocks $A_{k_1,l_1}, \ldots, A_{k_d,l_d}$. Indeed, we can choose at least $\lceil d/2 \rceil$ columns from different blocks of A_2 and complete the basis of $\mathbb{F}_2^d$ by selecting the remaining, i.e. at most $\lfloor d/2 \rfloor$, columns from different blocks of A_1. Let $A' = (a^i_{k_j, 2l_j - e_j})_{i,j=1}^d$ be the resulting matrix; since A' is nonsingular, simple row operations, such as addition and permutation, in matrix A (this corresponds to a change of the $\ker(L)$ basis) will lead to $A' = I$, that is A' becomes the identity matrix. □

Corollary 2. *With the notation of Theorem 3, let* $d = \dim \ker(L)$ *where mapping* L *is given by* (12)*. Then, we have that*

$$h_{f_1} + h_{f_2} - d \leq h_{f_1+f_2} \leq h_{f_1} + h_{f_2} - \lceil d/2 \rceil \tag{18}$$

whereas $\lambda^b_{f_1+f_2} = \lambda^{c_1}_{f_1} + \lambda^{c_2}_{f_2}$ *holds for* $\big(2^{2h_{f_1}+2h_{f_2}-2d+1} - 1\big)2^{2m-1} + 2^{m-1}$ *number of vectors* $b \preceq b'$*.*

Proof. It is easily seen that (18) can be directly obtained from the analysis following (16) in the proof of Theorem 3. Moreover, the number of times for which $\epsilon(b)$ is nonzero is equal to $2^{2m-1} - 2^{m-1}$ [30, p. 441]. Therefore $\lambda^b_{f_1+f_2} = \lambda^{c_1}_{f_1} + \lambda^{c_2}_{f_2}$ holds for $2^{2h_{f_1+f_2}} - 2^{2m-1} + 2^{m-1}$ vectors b. □

Remark 3. The choice of the first $2m$ elements of $\mathcal{E}^1_d$ to form $\widehat{\mathcal{E}}^1_d$ is one of the $\binom{d}{2m}$ possible alternatives; obviously, any of these choices is valid. Moreover, it is clear by the proof of Theorem 3 that if $h_f = h_{f_1} + h_{f_2} - d$, that is no new quadratic terms appear in function f, then we get $m = 0$, $\lambda^b_f = \lambda^{c_1}_{f_1} + \lambda^{c_2}_{f_2}$, and (17) becomes $c(\mathcal{E}^1_d) = \psi(a^i) + \zeta(a^i, b)$, since $b(\mathcal{E}^1_d)$ is the all-zero vector. This situation typically arises if functions f_1, f_2 have in common $\lfloor d/2 \rfloor$ products $g_{1,2i-1}g_{1,2i} = g_{2,2j-1}g_{2,2j}$, for some $1 \leq i \leq h_{f_1}$ and $1 \leq j \leq h_{f_2}$, in which case vectors $a^1, \ldots, a^d$ have weight 2. □

Theorem 4. *Let the Boolean functions* $f_1, \ldots, f_s \in \mathfrak{R}(2, n)$*, with* $s \geq 2$*, be given by* (7)*. Then, for some properly chosen fixed binary vector* b'*, of length* $2h_{f_1} + \cdots + 2h_{f_s}$ *and weight* $2h_{f_1+\cdots+f_s}$*, we have*

$$\lambda^b_{f_1+\cdots+f_s} = \lambda^{c_1}_{f_1} + \cdots + \lambda^{c_s}_{f_s} + \epsilon(b)\,, \quad \forall\ b \preceq b' \tag{19}$$

where $c_i \triangleq c_i(b) \in \mathbb{F}_2^{2h_{f_i}}$ *and* $\epsilon(b) \in \mathbb{F}_2$ *depend on vector* b*.*

Proof. Let us denote h_{f_i} by h_i for simplicity and define $f = f_1 + \cdots + f_s$. The validity of (19) can be readily established by recursive application of Theorem 3. The best affine approximations of function f are computed as $(\cdots((f_1 + f_2) + f_3)\cdots) + f_s$, i.e. at the ith step, $2 \leq i \leq s$, we have that

$$\lambda^{b_i}_{f_1+\cdots+f_i} = \lambda^{\tilde{c}_i(b_i)}_{f_1+\cdots+f_{i-1}} + \lambda^{\hat{c}_i(b_i)}_{f_i} + \hat{\epsilon}_i(b_i)\,, \quad \forall\ b_i \preceq b'_i \tag{20}$$

where vector b'_i has length $2H_i = 2(h_1 + \cdots + h_i)$ and weight $2h_{f_1+\cdots+f_i}$, with $b'_s = b'$ and $b_s = b$. Let us regard $\tilde{c}_i$, $\hat{c}_i$, and $\hat{\epsilon}_i$ as functions

$$\tilde{c}_i : \mathbb{F}_2^{2H_i} \mapsto \mathbb{F}_2^{2H_{i-1}}, \quad \hat{c}_i : \mathbb{F}_2^{2H_i} \mapsto \mathbb{F}_2^{2h_i}, \quad \text{and} \quad \hat{\epsilon}_i : \mathbb{F}_2^{2H_i} \mapsto \mathbb{F}_2$$

where $\tilde{c}_i$, $\hat{c}_i$ are defined by (17), and $\hat{\epsilon}_i(b_i) = \psi(b_i(\widehat{\mathcal{E}}^1_{d_i}))$. Moreover, let $\hat{c}_1$, $\hat{\epsilon}_1$ be the identity and zero functions respectively. Then, (20) leads to

$$\lambda^b_{f_1+\cdots+f_s} = \sum_{i=1}^{s} \left(\lambda_{f_i}^{(\hat{c}_i \circ \tilde{c}_{i+1} \circ \cdots \circ \tilde{c}_s)(b)} + (\hat{\epsilon}_i \circ \tilde{c}_{i+1} \circ \cdots \circ \tilde{c}_s)(b) \right), \quad \forall\ b \preceq b'$$

where "$\circ$" denotes the composition of functions. Direct comparison of the above expression with (19) gives that $\epsilon(b) = \epsilon_1(b) + \cdots + \epsilon_s(b)$, where we have $\epsilon_i(b) = (\hat{\epsilon}_i \circ \tilde{c}_{i+1} \circ \cdots \circ \tilde{c}_s)(b)$, and $c_i(b) = (\hat{c}_i \circ \tilde{c}_{i+1} \circ \cdots \circ \tilde{c}_s)(b)$, for $1 \leq i \leq s$. □

Note that since the same expression is obtained (up to a permutation of the coordinates of b) regardless the ordering of the functions f_i, we have $c_i(b) = c_i^\pi(b)$ and $\epsilon(b) = \epsilon^\pi(b) = \epsilon_1^\pi(b) + \cdots + \epsilon_s^\pi(b)$, where

$$c_i^\pi(b) = (\hat{c}_i^\pi \circ \tilde{c}_{i+1}^\pi \circ \cdots \circ \tilde{c}_s^\pi)(b) \quad \text{and} \quad \epsilon_i^\pi(b) = (\hat{\epsilon}_i^\pi \circ \tilde{c}_{i+1}^\pi \circ \cdots \circ \tilde{c}_s^\pi)(b)$$

for all $1 \leq i \leq s$, and permutations $\pi \in \mathcal{P}_s$ that correspond to the ordering of functions $(\cdots((f_{\pi_1} + f_{\pi_2}) + f_{\pi_3})\cdots) + f_{\pi_s}$. Results similar to those of Corollary 2 can also be proved in the general case. In particular, we can always find a vector b' of length $2h_{f_1} + \cdots + 2h_{f_s}$ and weight less than or equal to $2h_{f_1+\cdots+f_s}$ such that it holds $\epsilon_i^\pi(b) = 0$ for all $1 \leq i \leq s$, $b \preceq b'$, and permutations $\pi \in \mathcal{P}_s$. Then, this property implies that for arbitrary fixed vector $b \preceq b'$ we have

$$\lambda^b_{r_1 f_1+\cdots+r_s f_s} = r_1 \lambda^{c_1}_{f_1} + \cdots + r_s \lambda^{c_s}_{f_s}, \quad \forall\ r \in \mathbb{F}_2^s \tag{21}$$

with $r = (r_1, \ldots, r_s)$. The all-zero vector $b' = 0$ is an obvious solution to the problem; more precisely, we have $\lambda^0_{r_1 f_1+\cdots+r_s f_s} = r_1 g_{1,0} + \cdots + r_s g_{s,0}$. In order to obtain a wider set of solutions, let us consider mapping (12), which becomes $L(b) = \sum_{i=1}^{s} \sum_{j=1}^{2h_i} b_{i,j} g_{i,j}$ with $b = (b_1, \ldots, b_s)$, and let its kernel dimension be equal to d. We likewise define $\mathcal{E}_d$, $\mathcal{E}^0_d$, and $\mathcal{E}^1_d$ (in this case $1 \leq k_i \leq s$). The vector b', which is given by $b'_{i,j} = 0$ if $(i,j) \in \mathcal{E}^0_d \cup \mathcal{E}^1_d$ and $b'_{i,j} = 1$ otherwise, is another solution. This is due to the fact that all the functions $\epsilon_i^\pi(b)$ are actually evaluated at a subset of the coordinates of $b \preceq b'$; in particular, they are evaluated at $b(\widehat{\mathcal{E}}^1_d) \subseteq b(\mathcal{E}^1_d)$. The same holds for the functions $c_i^\pi(b)$ that only change the values of $b(\mathcal{E}^1_d)$. Hence, we have proved the following.

Proposition 3. *With the above notation, there always exist vectors $b \preceq b'$ such that $\lambda^b_{r_1 f_1+\cdots+r_s f_s} = r_1 \lambda^{c_1}_{f_1} + \cdots + r_s \lambda^{c_s}_{f_s}$ for all $r = (r_1, \ldots, r_s) \in \mathbb{F}_2^s$.*

The preceding result will play a prominent role to subsequently derive a compact formula, similar to (9), for the best quadratic approximations of cubic Boolean functions.

4 Results on Best Quadratic Approximations

In this section we develop a framework for efficiently computing the best quadratic approximation of Boolean functions whose algebraic degree is equal to 3. First, we need to introduce the following classification of cubic Boolean functions on n variables.

Definition 2. *The Boolean function $f \in \mathfrak{R}(3,n)$ is said to be a* class-m function *if, under all affine transformations $f'(x) = f(Ax+b)$ with non-singular $n \times n$ matrix A, m is the smallest positive integer such that there exists a set $\mathcal{J} = \{j_1, \ldots, j_m\}$ with $1 \leq j_1 < \cdots < j_m \leq n$ and the property that each cubic term of f' involves* at least one *variable with index in $\mathcal{J}$.*

It is clear from the above definition that $m \leq \lfloor n/3 \rfloor$, where the equality is obtained in the case the cubic part of f is comprised of $\lfloor n/3 \rfloor$ terms having pairwise no common variables. It is well-known (*see* e.g. [22], [30, p. 446]) that a cubic Boolean function $f \in \mathbb{B}_n$ such that $2^{n-3} \leq \mathrm{wt}(f) < 2^{n-2}$ can be transformed by an affine transformation into either

1. $x_1(x_2x_3 + \cdots + x_{2\mu}x_{2\mu+1})$, for $1 \leq \mu \leq \lfloor (n-1)/2 \rfloor$; or
2. $x_1x_2x_3 + x_4x_5x_6$, for $n \geq 6$.

Obviously, the above cases correspond to the class-1 and class-2 Boolean functions with $\mathcal{J} = \{1\}$ and $\mathcal{J} = \{1,4\}$ respectively. Note that the set $\mathcal{J}$ may not be unique for a given function, since many choices, out of the $\binom{n}{m}$ possible ones, may satisfy the conditions of Definition 2. The number of the equivalence classes is increased if $2^{n-3} \leq \mathrm{wt}(f) < 2^{n-2} + 2^{n-4}$ [23]; however they all are class-m, $1 \leq m \leq 3$, Boolean functions. Definition 2 implies that any cubic Boolean function belongs to exactly one class, due to the minimality of m. An important subset of class-m Boolean functions are the *separable class-m functions* whose cubic terms involve *exactly one* variable with index in $\mathcal{J}$ (as the ones presented above); their cubic part is equal to $c = \sum_{i=1}^{m} x_{j_i} q_i$, where the polynomials $q_i \in \mathbb{B}_{n-m}$ are quadratic and do not depend on variables with index in $\mathcal{J}$. As the equivalence classes of cubic functions found in [22, 23] are separable, this suggests that they comprise a large subset of cubic Boolean functions. Some properties that result from the preceding definitions are given below.

Lemma 2. *Let $f, g \in \mathfrak{R}(3,n)$ be cubic Boolean functions with the same cubic part. Then, $\mathcal{NQ}_f = \mathcal{NQ}_g$.*

Proof. By the definition of nonquadricity we get that for $\xi_f \in \mathcal{Q}_f$ it holds $\mathcal{NQ}_f = \mathrm{wt}(f + \xi_f) \leq \mathrm{wt}(f+h)$ for all $h \in \mathfrak{R}(2,n)$. Note that the function

$f+g$ is quadratic since f and g have the same cubic part. Therefore, by choosing $h=(f+g)+\xi_g$ we get that $\mathcal{NQ}_f \leq \mathrm{wt}(g+\xi_g)=\mathcal{NQ}_g$. Working similarly we may also derive $\mathcal{NQ}_g \leq \mathcal{NQ}_f$. Hence, $\mathcal{NQ}_f=\mathcal{NQ}_g$. □

Note that Lemma 2 is the natural extension of Lemma 1 in the quadratic approximation case, as it will be shown that only the cubic terms actually determine the best quadratic approximations of a cubic Boolean function. This is a well-known property, since adding a function of degree at most r to a function f, with $\deg(f)>r$, does not change $\mathcal{NL}_f^r$ (*see* e.g. [9]).

Proposition 4. *All class-m cubic Boolean functions $f\in\mathfrak{R}(3,n)$, with $\mathcal{J}=\{j_1,\ldots,j_m\}$, admit the following properties*

1. *Let $\mathcal{J}'\subset\mathcal{J}$ with cardinality k, $1\leq k\leq m-1$. From the decomposition $f=f_0\,\|_{\mathcal{J}'}\cdots\|_{\mathcal{J}'}\,f_{2^k-1}$ we have that all $f_i\in\mathbb{B}_{n-k}$ are class-$(m-k)$ cubic Boolean functions with the same cubic part;*
2. *Moreover, m is the least integer such that $f=f_0\,\|_{\mathcal{J}}\cdots\|_{\mathcal{J}}\,f_{2^m-1}$ with $\deg(f_i)<\deg(f)=3$ for all $0\leq i<2^m$.*

Proof. Without loss of generality assume that $\mathcal{J}'$ is comprised of the last k elements of the set $\mathcal{J}$, with $1\leq k\leq m-1$. We proceed by induction on the cardinality k of $\mathcal{J}'$. It is easily seen that Property 1 holds for $k=1$, since by $f=f_0\,\|_{j_m}\,f_1$ and the hypothesis, we conclude that $\deg(f_0)=3$ and its cubic part includes the cubic terms involving at least one variable with index in $\mathcal{J}\setminus\mathcal{J}'$; hence, f_0 is a class-$(m-1)$ cubic Boolean function. By Remark 1, it holds $f_1=f_0+f_1'$ with $\deg(f_1')<3$, and thus f_1 has the same cubic part with f_0. Next, assume that Property 1 holds for some k and $\mathcal{J}'=\{j_{m-k+1},\ldots,j_m\}$, $1\leq k<m-1$. The fact that it also holds for $k+1$ is established by the identity

$$\begin{aligned} f &= f_0\,\|_{\mathcal{J}'}\cdots\|_{\mathcal{J}'}\,f_{2^k-1}=(f_0'\,\|_{j_{m-k}}\,f_1')\,\|_{\mathcal{J}'}\cdots\|_{\mathcal{J}'}\,(f_{2^{k+1}-2}'\,\|_{j_{m-k}}\,f_{2^{k+1}-1}')\\ &= f_0'\,\|_{\{j_{m-k}\}\cup\mathcal{J}'}\cdots\|_{\{j_{m-k}\}\cup\mathcal{J}'}\,f_{2^{k+1}-1}' \end{aligned} \tag{22}$$

due to Definitions 1, 2, and the fact that $j_{m-k}<\min\mathcal{J}'$ (note that (22) would still hold, up to a re-ordering of the resulting sub-functions, if this was not true). Clearly, the sub-functions f_i (resp. f_i') include cubic terms involving at least one variable with index in $\mathcal{J}\setminus\mathcal{J}'$ (resp. $\mathcal{J}\setminus\{j_{m-k}\}\cup\mathcal{J}'$).

In order to prove Property 2 we need only consider (22) for $k=m-1$. From Property 1 we get that $f=f_0\,\|_{\mathcal{J}\setminus\{j_1\}}\cdots\|_{\mathcal{J}\setminus\{j_1\}}\,f_{2^{m-1}-1}$, where all f_i are class-1 cubic Boolean functions with the same cubic part (that of f_0). From (22) we have $f_i=f_{2i}'\,\|_{j_1}\,f_{2i+1}'$ and it is clear that both f_{2i}', f_{2i+1}' are quadratic. □

Consequently, Proposition 4 leads to an alternative definition of class-m cubic Boolean functions; that is, $m > 0$ is the least integer such that a proper choice of m variables leads to a decrease in the degree of the sub-functions obtained if mth order Shannon's expansion is performed with respect to these variables. Bearing in mind that many Boolean functions used in cryptography are constructed this way by functions with smaller degree, and the fact that many cryptographic criteria study properties of the sub-functions (e.g. propagation criteria of degree k and order m), we see that the classification imposed by Definition 2 is of high importance.

Lemma 3. *Let $f \in \mathfrak{R}(3,n)$ be a separable class-m function, with cubic part $c = \sum_{i=1}^{m} x_{j_i} q_i$, where $q_i \in \mathbb{B}_{n-m}$ is quadratic function not depending on variables with index in $\mathcal{J} = \{j_1, \ldots, j_m\}$. Then, from the decomposition $f = f_0 \,\|_{\mathcal{J}} \cdots \|_{\mathcal{J}}\, f_{2^m-1}$ we get that*

$$f_r = q + \langle r, p \rangle + l_r \,, \qquad 0 \le r < 2^m \tag{23}$$

for a quadratic $q \in \mathbb{B}_{n-m}$ and affine $l_r \in \mathbb{B}_{n-m}$ Boolean functions, where $r = (r_1, \ldots, r_m)$ is the binary representation of r, and $p = (q_1, \ldots, q_m)$.

Proof. The Boolean function is written as $f = c + q + l$, where c, q, and l is its cubic, quadratic, and linear part respectively. By hypothesis, f is a class-m cubic Boolean function, and therefore according to Definition 2 we necessarily have that $q_1, \ldots, q_m \neq 0$ and linearly independent. Indeed, let us assume that there exist $a_1, \ldots, a_m \in \mathbb{F}_2$, not all of them being zero, such that $a_1 q_1 + \cdots + a_m q_m = 0$; without loss of generality let $a_m = 1$. Therefore, we have $c = (x_{j_1} + a_1 x_{j_m}) q_1 + \cdots + (x_{j_{m-1}} + a_{m-1} x_{j_m}) q_{m-1}$, and there exists an invertible linear transformation (mapping $x_{j_i} + a_i x_{j_m}$ to x_{j_i}, for $1 \le i < m$, and all the remaining variables to themselves) such that f become class-$(m-1)$ cubic Boolean function—contradiction. The quadratic and linear parts of f can be similarly written as

$$q = \sum_{i=1}^{m-1} \sum_{k=i+1}^{m} x_{j_i} x_{j_k} \epsilon_{i,k} + \sum_{i=1}^{m} x_{j_i} l_i + q' \quad \text{and} \quad l = \sum_{i=1}^{m} x_{j_i} \epsilon_i + l' \tag{24}$$

for some quadratic $q' \in \mathbb{B}_{n-m}$ and linear functions $l', l_i \in \mathbb{B}_{n-m}$ that do not depend on $x_{j_1}, \ldots, x_{j_m}$. Let us next introduce the auxiliary functions

$$g^s = \left(\sum_{i=1}^{s} x_{j_i} q_i \right) + \left(\sum_{i=1}^{s-1} \sum_{k=i+1}^{s} x_{j_i} x_{j_k} \epsilon_{i,k} + \sum_{i=1}^{s} x_{j_i} l_i + q' \right) + \left(\sum_{i=1}^{s} x_{j_i} \epsilon_i + l' \right)$$

where the parentheses are used to indicate its cubic, quadratic, and linear parts respectively (note that $g^m = f$ whereas $g^0 = q' + l'$), and

$$h_i^s = \sum_{k=1}^{s} x_{j_k}\epsilon_{k,i} + \sum_{k=i+1}^{m} r_k\epsilon_{i,k}\,, \qquad 0 \le s < i \le m$$

where $r_k \in \mathbb{F}_2$. By applying Shannon's expansion formula recursively, as we did in Proposition 4, we obtain at the first step $f = f_0 \parallel_{j_m} f_1$, where $f_{r_m} = g^{m-1} + r_m(q_m + l_m + \epsilon_m + h_m^{m-1})$ for $r_m = 0, 1$. Further expansion of these sub-functions gives $f = (f_0 \parallel_{j_{m-1}} f_1) \parallel_{j_m} (f_2 \parallel_{j_{m-1}} f_3)$, where

$$f_r = g^{m-2} + \sum_{i=m-1}^{m} r_i\big(q_i + l_i + \epsilon_i + h_i^{m-2}\big)\,, \qquad 0 \le r < 4$$

and $r = r_{m-1} + 2r_m$ is the binary expansion of r. Clearly, if we continue this way, we obtain the decomposition $f = f_0 \parallel_{\mathcal{J}} \cdots \parallel_{\mathcal{J}} f_{2^m-1}$ after $m-2$ steps, which for all $0 \le r < 2^m$ leads to

$$f_r = q' + \sum_{i=1}^{m} r_i q_i + \left(l' + \sum_{i=1}^{m} r_i\big(l_i + \epsilon_i + \textstyle\sum_{k=i+1}^{m} r_k\epsilon_{i,k}\big)\right) \tag{25}$$

and $r = r_1 + 2r_2 + \cdots + 2^{m-1}r_m$ is the binary expansion of r. The claim is proved by noting that the expression inside the parentheses corresponds to l_r in (23), q' corresponds to q, and $\langle r, p\rangle = \sum_{i=1}^{m} r_i q_i$. □

In the sequel we assume that the quadratic Boolean function $q \in \mathbb{B}_{n-m}$ in (23), which is comprised of the quadratic terms of f not depending on the variables $x_{j_1}, \ldots, x_{j_m}$, does not belong to the linear space induced by $q_1, \ldots, q_m$. Otherwise, if we have $q = a_1q_1 + \cdots + a_mq_m$ for some $a_i \in \mathbb{F}_2$, then $c+q = (x_{j_1}+a_1)q_1 + \cdots + (x_{j_m}+a_m)q_m$ and there exists a translation mapping $x_{j_i} + a_i$ to x_{j_i}, $1 \le i \le m$, such that q is considered to be zero. The case of class-1 cubic Boolean functions is of particular interest, since then all cubic terms have one variable, say x_j, in common and $c = x_jq_j$. In the sequel, we determine their best quadratic approximations.

Theorem 5. *With the notation of Lemma 3, assume $f \in \mathbb{B}_n$ is a class-1 cubic function, where $f = (q+l_0) \parallel_j (q+q_j+l_1)$. Then, the best quadratic approximations of f have one of the following forms*

i. $\xi_f^0 = (q + l_0) \parallel_j (q + l_1 + \lambda_{q_j})$;
ii. $\xi_f^1 = (q + q_j + l_0 + \lambda_{q_j}) \parallel_j (q + q_j + l_1)$.

Proof. First, note that both ξ_f^0, ξ_f^1 (in the sequel, they are referred to as *form-i functions*, for $i = 0, 1$) are quadratic Boolean functions since their sub-functions have the same quadratic part. Next, assume that $\xi \in \mathbb{B}_n$ is a form-0 Boolean function. Then, $f + \xi = 0 \parallel_j (q_j + \lambda_{q_j})$, which in turn leads to

$$\mathrm{wt}(f + \xi) = \mathrm{wt}(q_j + \lambda_{q_j}) = \mathcal{NL}_{q_j}$$

from the definition of nonlinearity, and the fact that λ_{q_j} is a best affine approximation of $q_j \in \mathfrak{R}(2, n-1)$. Therefore, by Theorem 1 we have that $\mathrm{wt}(f + \xi) = 2^{n-2} - 2^{n-2-h_{q_j}}$. It is clear that the same result is obtained in the case of form-1 quadratic Boolean functions; hence, they all have the same distance from f.

Next, we show that the distance of any other quadratic function from f is greater than $2^{n-2} - 2^{n-2-h_{q_j}}$, therefore proving that $\mathcal{Q}_f$ consists of exactly the form-i, $i = 0, 1$, Boolean functions. Let us assume there exists some function $u \in \mathfrak{R}(2, n)$, not of these forms, such that

$$\mathrm{wt}(f + u) \leq 2^{n-2} - 2^{n-2-h_{q_j}} \,. \tag{26}$$

By Remark 1, it is easily seen that we have $u = u_0 \parallel_j u_1$ and $u_i = q' + l_i'$, where q' is the quadratic and l_0', l_1' are the linear parts of u_0, u_1. Note that by hypothesis $q' \neq q, q + q_j$, otherwise if e.g. $q' = q$ then u does not satisfy (26) unless it corresponds to a form-0 Boolean function. Indeed, let $q' = q$; then we get $\mathrm{wt}(f + u) = \mathrm{wt}(l_0 + l_0') + \mathrm{wt}(q_j + l_1 + l_1')$, which is greater than 2^{n-2} unless we set $l_0' = l_0$. Therefore, we have

$$\mathrm{wt}(f + u) = \mathrm{wt}(q_j + l_1 + l_1') \geq \mathrm{wt}(q_j + \lambda_{q_j}) = \mathcal{NL}_{q_j}$$

where equality holds if and only if we set $l_1' = l_1 + \lambda_{q_j}$. Hence, we get that $u = (q + l_0) \parallel_j (q + l_1 + \lambda_{q_j})$—contradiction. So, for $q' \neq q, q + q_j$ we get

$$\begin{aligned} \mathrm{wt}(f + u) &= \mathrm{wt}(q' + q + l_0' + l_0) + \mathrm{wt}(q' + q + q_j + l_1' + l_1) \\ &\geq \mathrm{wt}(q' + q + \lambda_{q'+q}) + \mathrm{wt}(q' + q + q_j + \lambda_{q'+q+q_j}) \\ &= 2^{n-1} - 2^{n-2-h_{q'+q}} - 2^{n-2-h_{q'+q+q_j}} \end{aligned}$$

where equality holds if and only if $l_i' + l_i \in \mathcal{A}_{q'+q+iq_j} \Leftrightarrow l_i' \in \mathcal{A}_{q'+q+iq_j+l_i}$ for $i = 0, 1$, according to Lemma 1. Even if q' could be chosen such that $h_{q'+q} = h_{q'+q+q_j} = 1$, we would have $\mathrm{wt}(f + u) = 2^{n-2} > 2^{n-2} - 2^{n-2-h_{q_j}}$ for all $1 \leq h_{q_j} \leq \lfloor (n-1)/2 \rfloor$. □

Corollary 3. *With the notation of Theorem 5, the nonquadricity of any class-*1 *function $f \in \mathfrak{R}(3, n)$ is equal to $\mathcal{NQ}_f = 2^{n-2} - 2^{n-2-h_{q_j}}$, for some $1 \leq h_{q_j} \leq \lfloor (n-1)/2 \rfloor$.*

The importance of Theorem 5 rests with the fact that it enables direct computation of the best quadratic approximation, of a particular subset of cubic Boolean functions on n variables, by determining the best affine approximation of quadratic Boolean functions on $n-1$ variables. To this end, Theorem 2 is applied to provide a direct solution.

Example 3. Let $f \in \mathbb{B}_5$ be the cubic function $f(x_1,\ldots,x_5) = x_1x_2x_4 + x_2x_3x_5 + x_1x_5 + x_2x_3 + x_3x_4 + x_3x_5 + x_3$. It is clear that f satisfies the conditions of Theorem 5, since $\mathcal{J} = \{2\}$. Thus, we write f as

$$f = (x_1x_5 + x_3x_4 + x_3x_5 + x_3) \parallel_2 (x_1x_4 + x_1x_5 + x_3x_4)$$

and proceed with the computation of a best affine approximation of the quadratic function $q_2 = x_1x_4 + x_3x_5$. By Theorem 2, $\lambda_{q_2} = 0$ is one of the solutions. Hence, we get the following best quadratic approximations

$$\xi_f^0 = x_1x_5 + x_2x_3 + x_3x_4 + x_3x_5 + x_3$$
$$\xi_f^1 = x_1x_4 + x_1x_5 + x_2x_3 + x_3x_4 + x_3$$

It is easily seen that $\mathrm{wt}(f + \xi_f^0) = \mathrm{wt}(f + \xi_f^1) = 6 = \mathcal{NQ}_f$, as in this case we have $h_{q_2} = 2$. □

Subsequently we develop the necessary background, by proving a series of results, in order to introduce the construction method of finding the best quadratic approximations of class-m cubic Boolean functions, $m > 1$.

Lemma 4. *Let $f \in \mathbb{B}_n$ be a Boolean function having the decomposition $f = f_0 \parallel_{\mathcal{J}} \cdots \parallel_{\mathcal{J}} f_{2^m-1}$, for some $m > 0$ and $\mathcal{J} = \{j_1,\ldots,j_m\}$, where $f_r \in \mathbb{B}_{n-m}$ do not depend on variables with index in $\mathcal{J}$. Then*

$$f = \sum_{c \in \mathbb{F}_2^m} \left(\sum_{r \preceq c} f_r \right) x_{j_1}^{c_1} \cdots x_{j_m}^{c_m} . \tag{27}$$

Proof. We proceed by induction on the cardinality m of $\mathcal{J}$. Clearly, (27) holds for $m = 1$ (*see* Remark 1) and $m = 2$, since then it is easily found that $f = f_0 + x_{j_1}(f_0 + f_1) + x_{j_2}(f_0 + f_2) + x_{j_1}x_{j_2}(f_0 + f_1 + f_2 + f_3)$. Let (27) hold for the set $\mathcal{J}' = \mathcal{J} \setminus \{j_m\}$ of cardinality $m-1$. Then, from

$$f = \left(f_0 \parallel_{\mathcal{J}'} \cdots \parallel_{\mathcal{J}'} f_{2^{m-1}-1}\right) \parallel_{j_m} \left(f_{2^{m-1}} \parallel_{\mathcal{J}'} \cdots \parallel_{\mathcal{J}'} f_{2^m-1}\right) = f_0' \parallel_{j_m} f_1'$$

we get that $f_i' = \sum_{c \in \mathbb{F}_2^{m-1}} \left(\sum_{r \preceq c} f_{r+i\,2^{m-1}}\right) x_{j_1}^{c_1} \cdots x_{j_{m-1}}^{c_{m-1}}$ by the induction hypothesis, for $i = 0, 1$. Since $f = \sum_{c_m \in \mathbb{F}_2} (f_0' + c_m f_1') x_{j_m}^{c_m}$, it is readily established that (27) also holds for the set $\mathcal{J}$ with cardinality m. □

Corollary 4. *With the notation of Lemma 4, let $f_r = q + l_r$ where q, l_r is its quadratic and linear part, $0 \le r < 2^m$. Then,* $\deg(f) = 2$ *if and only if any of the following equivalent conditions holds for all $c \in \mathbb{F}_2^m$*

i. $\sum_{r \preceq c} l_r = \epsilon_c$ *if* $\mathrm{wt}(c) = 2$, *and* $\sum_{r \preceq c} l_r = 0$ *if* $\mathrm{wt}(c) \ge 3$ *;*
ii. $\sum_{r \preceq c} l_{r+s} = \epsilon_c$ *for all* $s \in \mathbb{F}_2^m$ *and* $\mathrm{wt}(c) = 2$ *;*

where $\epsilon_c \in \mathbb{F}_2$ depends only on c, and $r + s = (r_1 + s_1, \ldots, r_m + s_m)$.

Proof. The first condition (in which $\{r : r \preceq c\}$ forms a $\mathrm{wt}(c)$-dimensional subspace of $\mathbb{F}_2^m$) is a direct result of Lemma 4, as for all nonzero $c \in \mathbb{F}_2^m$ the coefficient $\sum_{r \preceq c} f_r$ in (27) involves even number of summands, and therefore we get that $\sum_{r \preceq c} f_r = \sum_{r \preceq c} l_r$. To prove the equivalence of the two conditions we use the following property

$$\sum_{r \preceq c} l_r = \sum_{s \preceq d} \sum_{r \preceq c+d} l_{r+s} \,, \qquad \forall\, c, d \in \mathbb{F}_2^m : d \preceq c \tag{28}$$

but with d chosen to satisfy $\mathrm{wt}(d) = \mathrm{wt}(c) - 2$ (and $\mathrm{wt}(c+d) = 2$). Note that if $s \preceq c$, then $r + s \preceq c$ for all $r \preceq c$, and hence condition-ii (in which $\{r + s : r \preceq c\}$ forms a $\mathrm{wt}(c)$-dimensional flat of $\mathbb{F}_2^m$) includes the part of condition-i corresponding to $\mathrm{wt}(c) = 2$.

(i) $\Rightarrow$ (ii): Let us assume $\mathrm{wt}(c) = 3$ and set $c' = c + d$. Then, from (28) and condition-i we get $0 = \sum_{r \preceq c} l_r = \sum_{r \preceq c'} l_r + \sum_{r \preceq c'} l_{r+d}$, which leads to $\sum_{r \preceq c'} l_{r+d} = \epsilon_{c'}$ for $d \in \mathbb{F}_2^m$ with $\mathrm{wt}(d) = 1$. Suppose condition-ii holds for all $0 \le \mathrm{wt}(d) < w$, and let $c \in \mathbb{F}_2^m$ such that $\mathrm{wt}(c) = w + 2$. Then, by the induction hypothesis, condition-i, and (28) we have that

$$\sum_{r \preceq c} l_r = 0 \;\Rightarrow\; \sum_{r \preceq c'} l_{r+d} = \sum_{s \preceq d,\, s \neq d} \sum_{r \preceq c'} l_{r+s} = (2^w - 1)\epsilon_{c'} = \epsilon_{c'}$$

since $0 \le \mathrm{wt}(s) < w = \mathrm{wt}(d)$. Hence $\sum_{r \preceq c'} l_{r+d} = \epsilon_{c'}$ for all $d \in \mathbb{F}_2^m$.

(ii) $\Rightarrow$ (i): Direct consequence of (28), since then for all $\mathrm{wt}(c) \ge 3$ we have that $\sum_{r \preceq c} l_r = \sum_{s \preceq d} \sum_{r \preceq c+d} l_{r+s} = 2^{\mathrm{wt}(c)-2} \epsilon_{c+d} = 0$. □

Remark 4. A direct consequence of the conditions given in Corollary 4 is that all sums taken over flats of $\mathbb{F}_2^m$ with dimension $\mathrm{wt}(c) \ge 3$ vanish. This is proved by induction on $\mathrm{wt}(c)$ and by writing a $\mathrm{wt}(c)$-dimensional flat as the difference of two subspaces of dimensions $\mathrm{wt}(c)$ and $\mathrm{wt}(c) + 1$ respectively. □

Remark 5. It is clear that the family of affine functions $l_r \in \mathbb{B}_{n-m}$, which were introduced in (25), satisfy the conditions of Corollary 4. Indeed, for all $c \in \mathbb{F}_2^m$ such that $2 \leq \mathrm{wt}(c) = s \leq m$ we have

$$\sum_{r \preceq c} l_r = 2^s l' + 2^{s-1} \sum_{i=1}^{m} c_i \big(l'_i + \epsilon_i\big) + 2^{s-2} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} c_i c_j \epsilon_{i,j} \, .$$

Hence $\sum_{r \preceq c} l_r = \epsilon_{i,j}$ if $s = 2$ (where $c_i = c_j = 1$ and $c_k = 0$ for $k \neq i, j$), and zero for $s \geq 3$. Moreover, it is easily proved, by induction on s, that a family of 2^m affine functions $\{l_r : 0 \leq r < 2^m\}$ satisfies the conditions of Corollary 4 if and only if $l_r = l' + \sum_{i=1}^{m} r_i l'_i + \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} r_i r_j \delta_{i,j}$ for the affine functions $l' = l_0$, $l'_i = l_0 + l_{2^{i-1}}$, and $\delta_{i,j} \in \mathbb{F}_2$ (the sufficiency part has already been proved above). □

Next, we prove that the best quadratic approximations of a class-m cubic Boolean function, with $m \geq 2$, cannot be found recursively using the best quadratic approximations of the contained class-1 sub-functions.

Proposition 5. *With the notation of Lemma 3, let $f \in \mathbb{B}_n$ be a separable class-2 function with $\mathcal{J} = \{i, j\}$ and let $f = f_0 \parallel_j f_1$. Then, no pair of the functions $(\xi_{f_0}, \xi_{f_1}) \in \mathcal{Q}_{f_0} \times \mathcal{Q}_{f_1}$ has the same quadratic part.*

Proof. From the proof of Lemma 3, the Boolean function f is written as $f = (x_i q_i + x_j q_j) + (x_i x_j \epsilon_{i,j} + x_i l_i + x_j l_j + q') + (x_i \epsilon_i + x_j \epsilon_j + l')$, where the parentheses are used to indicate its cubic, quadratic, and linear parts respectively. According to Proposition 4, f_0, f_1 are class-1 cubic Boolean functions, and by Lemma 3 we get that

$$f_0 = (q' + l_0) \parallel_i (q' + q_i + l_1) \tag{29a}$$

$$f_1 = (q' + q_j + l_2) \parallel_i (q' + q_i + q_j + l_3) \tag{29b}$$

where $l_{r_i + 2r_j} = l' + r_i(l_i + \epsilon_i) + r_j(l_j + \epsilon_j) + r_i r_j \epsilon_{i,j}$ and $r_i, r_j \in \mathbb{F}_2$. Thus, by Theorem 5, the best quadratic approximations of f_0, f_1 are

$$\begin{aligned}
\xi^0_{f_0} &= q' + x_i(l_0 + l_1 + \lambda_{q_i}) + l_0 \, , \\
\xi^1_{f_0} &= q' + q_i + x_i(l_0 + l_1 + \lambda_{q_i}) + l_0 + \lambda_{q_i} \, , \\
\xi^0_{f_1} &= q' + q_j + x_i(l_2 + l_3 + \lambda_{q_i}) + l_2 \, , \\
\xi^1_{f_1} &= q' + q_i + q_j + x_i(l_2 + l_3 + \lambda_{q_i}) + l_2 + \lambda_{q_i} \, .
\end{aligned}$$

Even though $l_2 + l_3 = l_0 + l_1 + \epsilon_{i,j}$ by Remark 5 (and thus the quadratic terms of $\xi^0_{f_0}, \xi^1_{f_0}, \xi^0_{f_1}, \xi^1_{f_1}$ involving x_i coincide), the claim is established by the fact that we have $q_i, q_j \neq 0$ and $q_i \neq q_j$ (check the details of the proof of Lemma 3). □

The following result generalizes Theorem 5 to the case of class-m Boolean functions for any $m \geq 1$. From the subsequent analysis it becomes evident that class-1 functions constitute the most cryptographically weak class of Boolean functions in terms of nonquadricity. The class-m cubic Boolean functions attain high nonquadricity, for large values of m, whereas their security is also attributed to the fact that the difficulty of finding a set $\mathcal{J}$ such that all the 2^m sub-functions in $f = f_0 \,\|_{\mathcal{J}} \cdots \|_{\mathcal{J}} f_{2^m-1}$ are quadratic (in order to find their best quadratic approximations as shown next) grows exponentially with m for a fixed number of variables n. First, we need to prove the following lemma.

Lemma 5. *For all integers k and vectors $a = (a_1, \ldots, a_s) \in \mathbb{Z}^s$, $s \geq 1$, the expression $V_s^k(a) = \sum_{r \in \mathbb{F}_2^s} 2^{k-\langle r,a \rangle}$ is equal to*

$$V_s^k(a) = 2^{k-\langle 1_s,a \rangle} \prod_{i=1}^{s} \big(2^{a_i} + 1\big) \tag{30}$$

where $r = (r_1, \ldots, r_s)$ and $1_s = (1, \ldots, 1)$ of length s.

Proof. Note that (30) holds for $s = 1$, since $2^k + 2^{k-a_1} = 2^{k-a_1}(2^{a_1} + 1)$, and suppose it is valid for some $s \geq 1$, and all $a = (a_1, \ldots, a_s)$. Then, let $a' = (a, a_{s+1})$ for all integers a_{s+1}, and $r' = (r, r_{s+1})$ with $r_{s+1} \in \mathbb{F}_2$; the induction hypothesis leads to

$$\begin{aligned} V_{s+1}^k(a') &= \sum_{r' \in \mathbb{F}_2^{s+1}} 2^{k-\langle r,a \rangle - r_{s+1}a_{s+1}} = V_s^k(a) + V_s^{k-a_{s+1}}(a) \\ &= 2^{-a_{s+1}} V_s^k(a)\big(2^{a_{s+1}} + 1\big) = 2^{k-\langle 1_{s+1},a' \rangle} \prod_{i=1}^{s+1} \big(2^{a_i} + 1\big) \end{aligned}$$

for all $a' = (a_1, \ldots, a_{s+1}) \in \mathbb{Z}^{s+1}$, since $V_s^{k-a_{s+1}}(a) = 2^{-a_{s+1}} V_s^k(a)$, hence concluding our proof. □

Theorem 6. *With the notation of Lemma 3, assume $f \in \mathbb{B}_n$ is a class-m cubic function, and let $q_i \in \mathbb{B}_{n-m}$ be given by (7). If all linear functions in $\bigcup_{i=1}^{m} \{ g_{i,1}, \ldots, g_{i,2h_{q_i}} \}$ are linearly independent, then the best quadratic approximations of f have one of the following forms*

$$\xi_f^s = \xi_{f,0}^s \,\|_{\mathcal{J}} \cdots \|_{\mathcal{J}} \, \xi_{f,2^m-1}^s \,, \qquad 0 \leq s < 2^m \tag{31}$$

where $\xi_{f,r}^s = q + \langle s, p \rangle + l_r + \lambda_{\langle r+s,p \rangle}$ and $r + s = (r_1 + s_1, \ldots, r_m + s_m)$, for $0 \leq r < 2^m$.

Proof. First, note that all the form-s sub-functions in (31) have the same quadratic part $q + \langle s, p\rangle$, which is necessary, but not sufficient, in order for these Boolean functions to be quadratic. According to Lemma 4 and Corollary 4 they need also satisfy for all $c, s \in \mathbb{F}_2^m$ the following condition

$$\sum_{r \preceq c} \big(l_r + \lambda_{\langle r+s,p\rangle}\big) = \begin{cases} \delta_c, & \text{if } \operatorname{wt}(c) = 2\,, \\ 0, & \text{if } \operatorname{wt}(c) \geq 3\,, \end{cases} \tag{32}$$

for some constants $\delta_c \in \mathbb{F}_2$ which depend on c. However, from Remark 5 we have $\sum_{r \preceq c} l_r = \epsilon_c$ for some $\epsilon_c \in \mathbb{F}_2$ if $\operatorname{wt}(c) = 2$, and zero if $\operatorname{wt}(c) \geq 3$. By also taking into account Remark 4, we conclude that ξ_f^s are quadratic functions if and only if it holds $\sum_{r \preceq c} \lambda_{\langle r+s,p\rangle} = \delta_c + \epsilon_c$ for $\operatorname{wt}(c) = 2$ and all $s \in \mathbb{F}_2^m$. From Proposition 3, we see that there always exists a proper choice of best affine approximations $\lambda_{q_1}, \ldots, \lambda_{q_m}$ such that

$$\delta_c + \epsilon_c = \sum_{r \preceq c} \lambda_{\langle r+s,p\rangle} = \sum_{r \preceq c} \left(\sum_{i=1}^{m} (r_i + s_i) \lambda_{q_i} \right) = \sum_{i=1}^{m} \left(\sum_{r \preceq c} (r_i + s_i) \lambda_{q_i} \right) = 0$$

since $\sum_{r \preceq c} (r_i + s_i)\lambda_{q_i}$ equals $4s_i\lambda_{q_i}$ if $c_i = 0$, and $2(2s_i + 1)\lambda_{q_i}$ otherwise. Therefore (32) is satisfied with $\delta_c = \epsilon_c$, and ξ_f^s are quadratic functions for all $s \in \mathbb{F}_2^m$.

Next, let us suppose that $\xi \in \mathbb{B}_n$ is some form-s Boolean function. Then, the sth sub-function of $f + \xi$ is identically zero since the respective sub-functions of f, ξ coincide due to the construction of ξ, while its rth sub-function for $r \neq s$ equals $\langle r + s, p\rangle + \lambda_{\langle r+s,p\rangle}$, leading to

$$\operatorname{wt}(f + \xi) = \sum_{r \in \mathbb{F}_2^m \setminus \{s\}} \operatorname{wt}(\langle r + s, p\rangle + \lambda_{\langle r+s,p\rangle}) = \sum_{r \in \mathbb{F}_2^m \setminus \{0\}} \operatorname{wt}(\langle r, p\rangle + \lambda_{\langle r,p\rangle})$$

due to the fact that adding some fixed $s \in \mathbb{F}_2^m$ to all vectors in $\mathbb{F}_2^m$ results into a permutation of its elements. Thus, by the definition of nonlinearity, Theorem 1, Lemma 5, and the fact that $\mathcal{NL}_0 = h_0 = 0$ by convention, we have that $1 \leq h_{\langle r,p\rangle} \leq \lfloor (n-m)/2 \rfloor$ for $r \neq 0$ and

$$\begin{aligned} \operatorname{wt}(f + \xi) &= 2^{n-1} - \sum_{r \in \mathbb{F}_2^m} 2^{n-m-1-h_{\langle r,p\rangle}} = 2^{n-1} - V_m^{n-m-1}(h_{q_1}, \ldots, h_{q_m}) \\ &= 2^{n-1} - 2^{n-m-1-(h_{q_1} + \cdots + h_{q_m})} \prod_{i=1}^{m} \big(2^{h_{q_i}} + 1\big) \end{aligned} \tag{33}$$

since $\bigcup_{i=1}^{m} \{\, g_{i,1}, \ldots, g_{i,2h_{q_i}} \}$ are linearly independent by hypothesis, and thus $h_{\langle r,p\rangle} = h_{r_1 q_1 + \cdots + r_m q_m} = r_1 h_{q_1} + \cdots + r_m h_{q_m}$ (*see* Theorem 3 proof).

Since the above expression is independent of s, it is clear that all quadratic Boolean functions given by (31), for $0 \leq s < 2^m$, have the same distance from the function f.

Next, we show that the distance of any other quadratic function from f is greater than $\mathrm{wt}(f+\xi)$, therefore proving that $\mathcal{Q}_f$ consists of exactly the form-s, for $0 \leq s < 2^m$, Boolean functions. Let us assume there exists a function $u \in \mathfrak{R}(2,n)$, which does not coincide with a form-s function, and $u = (q' + l'_0) \parallel_{\mathcal{J}} \cdots \parallel_{\mathcal{J}} (q' + l'_{2^m-1})$, where l'_r satisfy the conditions of Corollary 4. By hypothesis, we necessarily have that $q' \neq q + \langle s, p\rangle$ for all $0 \leq s < 2^m$, or equivalently $\tilde{q} \triangleq q' + q, q_1, \ldots, q_m$ are linearly independent; otherwise u would correspond to a form-s quadratic function (arguments similar to those given in the proof of Theorem 5 apply). Then, by setting $\tilde{l}_r = l'_r + l_r$ we likewise find that

$$\begin{aligned} \mathrm{wt}(f+u) &= \sum_{r \in \mathbb{F}_2^m} \mathrm{wt}(\tilde{q} + \langle r, p\rangle + \tilde{l}_r) \geq \sum_{r \in \mathbb{F}_2^m} \mathrm{wt}(\tilde{q} + \langle r, p\rangle + \lambda_{\tilde{q}+\langle r,p\rangle}) \\ &= 2^{n-1} - \sum_{r \in \mathbb{F}_2^m} 2^{n-m-1-h_{\tilde{q}+\langle r,p\rangle}} \end{aligned} \tag{34}$$

where equality holds if and only if $\tilde{l}_r \in \mathcal{A}_{\tilde{q}+\langle r,p\rangle}$, i.e. $\tilde{l}_r = \lambda_{\tilde{q}+\langle r,p\rangle}$, for all vectors $r \in \mathbb{F}_2^m$, according to Lemma 1. In order to minimize the weight of $f+u$, the quadratic function $\tilde{q}$ should be chosen such that all $h_{\tilde{q}+\langle r,p\rangle}$ take their minimum possible value. From the fact that $\tilde{q} + \langle r, p\rangle \neq 0$, we necessarily have that $h_{\tilde{q}+\langle r,p\rangle} \geq 1$ for all $r \in \mathbb{F}_2^m$; however, not all $h_{\tilde{q}+\langle r,p\rangle}$ can simultaneously be made equal to 1 if $m > 1$. Function $\tilde{q}$ is written as $\tilde{q} = \tilde{g}_0 + \sum_{j=1}^{h_{\tilde{q}}} \tilde{g}_{2j-1}\, \tilde{g}_{2j}$, where the linear part $\tilde{g}_0$ is obtained by applying Dickson's theorem on $\tilde{q}$.

Define d_i as the number of $\tilde{g}_j$ that are not linearly independent from $g_{i,j}$ of q_i, or that $\tilde{q}, q_i$ have in common, $1 \leq i \leq m$. Furthermore, let e_i be the number of products $\tilde{g}_{2j-1}\tilde{g}_{2j}$ shared by $\tilde{q}, q_i$. It is easily seen that we have $0 \leq d_i \leq 2\min\{h_{\tilde{q}}, h_{q_i}\}$ and $0 \leq e_i \leq \lfloor d_i/2 \rfloor$. From Corollary 2, we get $h_{\tilde{q}+q_i} \geq h_{\tilde{q}} + h_{q_i} - d_i$, whereas Remark 3 implies that the lower bound is always attained if d_i is even and $e_i = d_i/2$. Hence, the weight of $f+u$ can be minimized by letting $\tilde{q}$ have common products $\tilde{g}_{2j-1}\tilde{g}_{2j}$ with the quadratic functions $q_1, \ldots, q_m$. Since from hypothesis all functions in the set $\bigcup_{i=1}^{m} \{\, g_{i,1}, \ldots, g_{i,2h_{q_i}} \,\}$ are linearly independent, we get that

$$h_{\tilde{q}+\langle r,p\rangle} = h_{\tilde{q}} + \sum_{i=1}^{m} r_i\big(h_{q_i} - 2e_i\big) \quad \text{and} \quad 0 \leq \sum_{i=1}^{m} r_i e_i \leq \min\left\{ h_{\tilde{q}}, \sum_{i=1}^{m} r_i h_{q_i} \right\}.$$

From (34), Lemma 5, and the above relations, we conclude that $\mathrm{wt}(f+u)$ is greater than or equal to $2^{n-1} - 2^{-h_{\tilde{q}}} V_m^{n-m-1}(h_{q_1} - 2e_1, \ldots, h_{q_m} - 2e_m)$, depending on the parameters $e_1, \ldots, e_m$. Comparison with (33) gives

$$\begin{aligned} \mathrm{wt}(f+u) < \mathrm{wt}(f+\xi) &\Leftrightarrow 2^{h_{\tilde{q}}} \frac{V_m^{n-m-1}(h_{q_1}, \ldots, h_{q_m})}{V_m^{n-m-1}(h_{q_1} - 2e_1, \ldots, h_{q_m} - 2e_m)} < 1 \\ &\Leftrightarrow 2^{h_{\tilde{q}} - (e_1 + \cdots + e_m)} \prod_{i=1}^{m} \frac{2^{h_{q_i}} + 1}{2^{h_{q_i} - e_i} + 2^{e_i}} < 1 \,. \end{aligned} \tag{35}$$

Since it holds $0 \le e_i \le \min\{h_{\tilde{q}}, h_{q_i}\}$, all terms $(2^{h_{q_i}} + 1)(2^{h_{q_i} - e_i} + 2^{e_i})^{-1}$ in (35) are greater than or equal to 1, where equality is attained if either $e_i = 0$ or $e_i = h_{q_i} = \min\{h_{\tilde{q}}, h_{q_i}\}$; the latter case is valid for all $1 \le i \le m$ only if $h_{\tilde{q}} \ge \max\{h_{q_1}, \ldots, h_{q_m}\}$. Moreover, $e_i = 0$ implies that $\tilde{q}$ does not have common products with q_i, whereas $e_i = h_{q_i}$ that $\tilde{q}$ is written as the sum of q_i and another quadratic function. Since by hypothesis $\tilde{q} \neq \langle r, p \rangle$, we either have $0 < e_i < \min\{h_{\tilde{q}}, h_{q_i}\}$ for some $1 \le i \le m$, or that $\tilde{q}$ has a product whose functions do not depend on $\bigcup_{i=1}^{m} \{ g_{i,1}, \ldots, g_{i,2h_{q_i}} \}$, hence $2^{h_{\tilde{q}} - (e_1 + \cdots + e_m)} > 1$, due to $0 < e_1 + \cdots + e_m < \min\{h_{\tilde{q}}, h_{q_1} + \cdots + h_{q_m}\}$. Therefore, in any case we get $\mathrm{wt}(f+u) > \mathrm{wt}(f+\xi)$. □

In Theorem 6, we assumed that the functions $\bigcup_{i=1}^{m} \{ g_{i,1}, \ldots, g_{i,2h_{q_i}} \}$ are linearly independent. Since $h_{q_i} \ge 1$ for all $1 \le i \le m$, and the fact that it must also hold $h_{q_1} + \cdots + h_{q_m} = h_{q_1 + \cdots + q_m} \le \lfloor (n-m)/2 \rfloor$, we see that $h_{q_i} \le \lfloor (n-3m)/2 \rfloor + 1$.

Corollary 5. *With the notation of Theorem 6, the nonquadricity of any separable class-m cubic function $f \in \mathfrak{R}(3,n)$ is equal to*

$$\mathcal{NQ}_f = 2^{n-1} - 2^{n-m-1} \prod_{i=1}^{m} \left(1 + 2^{-h_{q_i}}\right) \tag{36}$$

for some $1 \le h_{q_i} \le \lfloor (n-3m)/2 \rfloor + 1$.

Let $P = (p_{i,j})_{i,j=1}^{m}$ be an $m \times m$ invertible matrix, $Q = (q_1, \ldots, q_m)$, and assume the vector $Q' = (q'_1, \ldots, q'_m)$ is given by $Q' = QP$. Then, from the independence of $\bigcup_{i=1}^{m} \{ g_{i,1}, \ldots, g_{i,2h_{q_i}} \}$ we get $h_{q'_j} = \sum_{i=1}^{m} p_{i,j} h_{q_i}$, and

$$h_{r_1 q'_1 + \cdots + r_m q'_m} = \sum_{i=1}^{m} \left(\sum_{j=1}^{m} r_j p_{i,j} \right) h_{q_i} = \sum_{i=1}^{m} r'_i h_{q_i} = h_{r'_1 q_1 + \cdots + r'_m q_m}$$

where $r'_i = \sum_{j=1}^{m} r_j p_{i,j} \bmod 2$, for $1 \le i \le m$. Hence, the results obtained in Theorem 6 and Corollary 5 would still hold in this case if we replace

$h_{q_i'}$ with the ith element of vector $Q'P^{-1}$. By considering the fact that all separable class-m cubic Boolean functions probably satisfy the conditions of Theorem 6 (*see* discussion at the beginning of the section and [22, 23]), it seems safe to assume that the applicability of the above results is more general than currently stated.

Theorem 7. *The covering radius of* $\mathfrak{R}(2,n)$ *in* $\mathfrak{R}(3,n)$ *admits the lower bound* $\rho_3(2,n) \geq 2^{n-1} - \frac{1}{2}6^{n/3}$*, corresponding to the nonquadricity of the separable class-*$\lfloor n/3 \rfloor$ *cubic Boolean functions.*

Proof. To derive the lower bound we need to determine the class of cubic Boolean functions that achieve the highest nonquadricity. By Corollary 5 we have that the nonquadricity of a cubic function f is maximized if and only if the product $\prod_{i=1}^{m}\big(1/2+1/2^{h_{q_i}+1}\big)$ depending on $m, h_{q_1}, \ldots, h_{q_m}$ is minimized. Since each product term is an integer less than 1, the number of terms m should be sufficiently large. However, the constraints on the values taken by h_{q_i} need also be considered. Let H be the set of distinct integers from $h_{q_1}+1, \ldots, h_{q_m}+1$, and suppose $a-r, a+r \in H$ for some $a > r+1 > 1$. Then, it is easily verified that we have the property

$$\left(\frac{1}{2}+\frac{1}{2^{a-r}}\right)\left(\frac{1}{2}+\frac{1}{2^{a+r}}\right) > \cdots > \left(\frac{1}{2}+\frac{1}{2^{a-1}}\right)\left(\frac{1}{2}+\frac{1}{2^{a+1}}\right) > \left(\frac{1}{2}+\frac{1}{2^{a}}\right)^2$$

from which we derive that $\max_{a\in H}\{a\} - \min_{a\in H}\{a\}$, and the cardinality of H, should be relatively small. Moreover, by noting that the sequence $\{(1/2+1/2^a)^i\}_{i\geq 0}$ is purely decreasing for any $a \in H$ (since then $a \geq 2$), we conclude that the highest possible nonquadricity achieved by separable class-m cubic Boolean functions, by Corollary 5, is given by

$$\begin{aligned}\max\nolimits_{\text{class-}m}\big\{\mathcal{NQ}\big\} &= 2^{n-1} - 2^{n-1}\left(\frac{1}{2}+\frac{1}{2^{a_m+2}}\right)^{b_m}\left(\frac{1}{2}+\frac{1}{2^{a_m+1}}\right)^{m-b_m} \\ &= 2^{n-1} - 2^{n-1}\left(\frac{2^{a_m+1}+1}{2^{a_m+1}+2}\right)^{b_m}\left(\frac{1}{2}+\frac{1}{2^{a_m+1}}\right)^{m} \qquad (37)\end{aligned}$$

where $a_m = \lfloor (n-m)/2m \rfloor$ and $b_m = \lfloor (n-m)/2 \rfloor \bmod m$, as a result of letting b_m functions q_i have $h_{q_i} = a_m + 1$, and the remaining $m - b_m$ have $h_{q_i} = a_m$. It is clear from (37) that for small values of m, the integer a_m is large and therefore the contribution of $(2^{a_m+1}+1)(2^{a_m+1}+2)^{-1} \approx 1$ is negligible (b_m is also small). Hence, the maximum nonquadricity attained by class-m cubic Boolean functions grows with $m \leq \lfloor n/3 \rfloor$. If $m = \lfloor n/3 \rfloor$, then we have $a_m = 1$, $b_m = \lfloor (n \bmod 3)/2 \rfloor$, and (37) becomes

$$\mathcal{NQ}_f = 2^{n-1} - 2^{\lceil (n \bmod 3)/2 \rceil - 1}\big(\tfrac{5}{3}\big)^{\lfloor (n \bmod 3)/2 \rfloor}6^{\lfloor n/3 \rfloor} = 2^{n-1} - b_n\tfrac{1}{2}6^{n/3}$$

Table 1. The maximum possible nonquadricity attained by the class-1 and class-$\lfloor n/3 \rfloor$ separable cubic Boolean functions in $\mathbb{B}_n$, for $3 \leq n \leq 27$, as computed by Theorem 7.

$\boldsymbol{n}$	3	6	9	12	15	18	21	24	27
class-1	1	12	120	992	8128	65280	523776	4192256	33550336
class-$\lfloor n/3 \rfloor$	1	14	148	1400	12496	107744	908608	7548800	62070016

where the term b_n equals 1 if $n \equiv 0 \pmod 3$, $(4/3)^{1/3}$ if $n \equiv 1 \pmod 3$, and $(250/243)^{1/3}$ if $n \equiv 2 \pmod 3$. Therefore, in all cases we have $b_n \approx 1$. The fact that we have only considered cubic Boolean functions satisfying the conditions of Theorem 6, leads to the lower bound. □

As seen from Theorem 7, class-$\lfloor n/3 \rfloor$ cubic functions achieve the highest possible nonquadricity among all separable class-m Boolean functions, for $1 \leq m \leq \lfloor n/3 \rfloor$, which satisfy the conditions of Theorem 6. As their cubic part is equivalent (under some transformation of variables $y = xR$) to

$$\sum_{i=1}^{\lfloor \frac{n}{3} \rfloor - 1} y_{3i-2}\, y_{3i-1}\, y_{3i} + y_{3\lfloor \frac{n}{3} \rfloor - 2}\Big(y_{3\lfloor \frac{n}{3} \rfloor - 1}\, y_{3\lfloor \frac{n}{3} \rfloor} + a\, y_{3\lfloor \frac{n}{3} \rfloor + 1}\, y_{3\lfloor \frac{n}{3} \rfloor + 2} \Big)$$

where $a = 1$ if $n \equiv 2 \pmod 3$ and zero otherwise, they can be considered as a natural extension of bent functions (they have similar representation and the maximum possible distance from all functions of degree one less) to the best quadratic approximation case. Their nonquadricity is depicted in Table 1.

Proposition 6. *With the notation of Theorem 6, the best quadratic approximations ξ_f^s of the class-m cubic Boolean function f are given by*

$$\xi_f^s = f + \sum_{i=1}^{m} (x_{j_i} + s_i)(q_i + \lambda_{q_i}) \,, \qquad 0 \leq s < 2^m \,. \tag{38}$$

Proof. From the proof of Theorem 6 and Definition 1, we have that for all $s \in \mathbb{F}_2^m$ the best quadratic approximation ξ_f^s of the class-m cubic function f, with cubic part $c = \sum_{i=1}^m x_{j_i} q_i$, is such that

$$\begin{aligned} \xi_f^s + f &= \left(\xi_{f,0}^s + f_0\right) \parallel_{\mathcal{J}} \cdots \parallel_{\mathcal{J}} \left(\xi_{f,2^m-1}^s + f_{2^m-1}\right) \\ &= \sum_{r \in \mathbb{F}_2^m} (x_{j_1} + \bar{r}_1) \cdots (x_{j_m} + \bar{r}_m) \sum_{i=1}^{m} (r_i + s_i)(q_i + \lambda_{q_i}) \end{aligned}$$

due to the linear independence of the functions in $\bigcup_{i=1}^m \{\, g_{i,1}, \ldots, g_{i,2h_{q_i}} \}$ and the fact that we may write $\lambda_{r_1 q_1 + \cdots + r_m q_m} = r_1 \lambda_{q_1} + \cdots + r_m \lambda_{q_m}$, for

all $r \in \mathbb{F}_2^m$. By writing the above expression as the sum of those terms for which $r_m = 0$ and those for $r_m = 1$, then simple calculations give

$$\begin{aligned}\xi_f^s + f = &\sum_{r \in \mathbb{F}_2^{m-1}} (x_{j_1} + \bar{r}_1) \cdots (x_{j_{m-1}} + \bar{r}_{m-1}) \sum_{i=1}^{m-1} (r_i + s_i)(q_i + \lambda_{q_i}) \\ &+ (x_{j_m} + s_m)(q_m + \lambda_{q_m}) \sum_{r \in \mathbb{F}_2^{m-1}} (x_{j_1} + \bar{r}_1) \cdots (x_{j_{m-1}} + \bar{r}_{m-1}) \\ = &\sum_{r \in \mathbb{F}_2^{m-1}} (x_{j_1} + \bar{r}_1) \cdots (x_{j_{m-1}} + \bar{r}_{m-1}) \sum_{i=1}^{m-1} (r_i + s_i)(q_i + \lambda_{q_i}) \\ &+ (x_{j_m} + s_m)(q_m + \lambda_{q_m})\end{aligned}$$

since it is easily seen that $\sum_r (x_{j_1} + \bar{r}_1) \cdots (x_{j_{m-1}} + \bar{r}_{m-1}) = 1$ corresponds to the constant all-one Boolean function. Clearly, repeated applications of the above steps will lead to (38). □

Given a class-m cubic Boolean function $f = \sum_{i=1}^m x_{j_i} q_i + q + l$, where q, l are its quadratic and linear parts respectively, and $q_1, \ldots, q_m$ satisfy the conditions of Theorem 6, its best quadratic approximations are directly computed by means of Proposition 6 as follows

$$\xi_f^s = \left(q + \sum_{i=1}^m \big(s_i q_i + x_{j_i} \lambda_{q_i} \big) \right) + \left(l + \sum_{i=1}^m s_i \lambda_{q_i} \right), \qquad 0 \le s < 2^m$$

where the parentheses indicate its quadratic and linear part respectively. The number of the best quadratic approximations depends on the number of the best affine approximations of functions $q_1, \ldots, q_m$. It is interesting to note the similarity of (9) and (38), i.e. in both cases we alter the highest degree terms with properly chosen functions of lower degree.

Example 4. Let $f \in \mathbb{B}_8$ be the class-2 Boolean function $f(x_1, \ldots, x_8) = (x_1 + x_3)(x_2 + x_7)(x_3 + x_5) + (x_4 + x_7)\big(x_5(x_6 + x_8) + (x_7 + x_8)x_8\big)$. It is easily seen that it satisfies the conditions of Theorem 6, and therefore its best quadratic approximations (from Proposition 6) are given by

$$\xi_f = s_1 q_1 + s_2 q_2 + (x_1 + x_3 + s_1)\lambda_{q_1} + (x_4 + x_7 + s_2)\lambda_{q_2}\,, \quad s_i \in \mathbb{F}_2$$

where $q_1 = (x_2 + x_7)(x_3 + x_5)$ and $q_2 = x_5(x_6 + x_8) + (x_7 + x_8)x_8$. From Section 3 we know that the best affine approximations of q_1, q_2 are

$$\lambda_{q_1} = a_1(x_2 + x_7) + a_2(x_3 + x_5) + a_1 a_2\,, \qquad a_i \in \mathbb{F}_2\,,$$

$$\lambda_{q_2} = b_1x_5 + b_2(x_6 + x_8) + b_3(x_7 + x_8) + b_4x_8 + b_1b_2 + b_3b_4\ , \quad b_i \in \mathbb{F}_2\ .$$

By Corollary 5 we see that its nonquadricity equals $\mathcal{NQ}_f = 2^7 - 2^2 3 \cdot 5 = 68$, which is the maximum possible since f is a class-$\lfloor 8/3 \rfloor$ function. □

The common characteristic of the cubic Boolean functions studied above is that their highest degree terms present common variables in a way that allows the efficient computation of their best quadratic approximations. These Boolean functions, called separable (the notion is readily extended to algebraic degrees greater than 3), have a particular structure that is undesirable in most cryptographic applications since it can be the source of many cryptanalytic attacks exploiting the existence of good low order approximations, e.g. linear cryptanalysis. Boolean functions that do not exhibit this structure are called *inseparable* and are known to exist [23]: e.g. $f = x_1x_2x_3 + x_4(x_1x_5 + x_2x_6 + x_3x_7 + x_8x_9 + \cdots + x_{2\mu}x_{2\mu+1})$ with $4 \leq \mu \leq \lfloor (n-1)/2 \rfloor$, which is an inseparable class-2 function. However, it is not known whether there always exists some m such that any cubic function is equivalent (under some affine transformation) to a separable class-m Boolean function. If this is true, then clearly our results cover the entire space of cubic Boolean functions.

5 Conclusions

This paper studied the problem of finding best low order approximations of Boolean functions. In particular, explicit formulas have been given for directly computing all best affine approximations of a quadratic function without use of the Walsh-Hadamard transform, as well as, for determining the best quadratic approximations of a separable cubic Boolean function. In correspondence with the nonlinearity, the notion of nonquadricity was introduced as the minimum distance from all quadratic functions, and classes of cubic functions that attain maximum nonquadricity were also identified. Due to the efficiency of low order approximation attacks, it is important to find Boolean functions that achieve maximum nonlinearity and nonquadricity, or to perform an in-depth analysis of highly nonlinear function constructions. Research in progress is focused on extending the results obtained to inseparable cubic functions and Boolean functions of higher degree. Furthermore, possible trade-offs between the nonquadricity and other cryptographic measures, like algebraic immunity, are currently studied.

References

1. Berbain, C., Billet, O., Canteaut, A., et. al.: DECIM - a new stream cipher for hardware applications. eSTREAM, ECRYPT Stream Cipher Project, Report 2005/004 (2005) http://www.ecrypt.eu.org/stream.
2. Borissov, Y., Braeken, A., Nikova, S., Preneel, B.: On the covering radii of binary ReedMuller codes in the set of resilient boolean functions. IEEE Transactions on Information Theory **51** (2005) 1182–1189.
3. Carlet, C.: Partially-bent functions. Designs, Codes and Cryptography **3** (1993) 135–145.
4. Carlet, C.: Two new classes of bent functions. Advances in Cryptology - Eurocrypt '93 (Lecture Notes in Computer Science, Springer-Verlag) **765** (1994) 77–101.
5. Carlet, C., Mesnager, S.: Improving the upper bounds on the covering radii of binary Reed-Muller codes. IEEE Transactions on Information Theory **53** (2007) 162–173.
6. Carlet, C.: Concatenating indicators of flats for designing cryptographic functions. Designs, Codes and Cryptography **36** (2005) 189–202.
7. Carlet, C., Dalai, D., Gupta, K., Maitra, S.: Algebraic immunity for cryptographically significant Boolean functions: analysis and construction. IEEE Transactions on Information Theory **52** (2006) 3105–3121.
8. Carlet, C.: On the higher order nonlinearities of algebraic immune functions. Advances in Cryptology - Crypto '06 (Lecture Notes in Computer Science, Springer-Verlag) **4117** (2006) 584–601.
9. Carlet, C.: Recursive lower bounds on the nonlinearity profile of Boolean functions and their applications. Cryptology ePrint Archive, Report 2006/459 (2006) http://eprint.iacr.org.
10. Charpin, P.: Normal boolean functions. Journal of Complexity **20** (2004) 245–265.
11. Charpin, P., Pasalic, E., Tavernier, C.: On bent and semi-bent quadratic boolean functions. IEEE Transactions on Information Theory **51** (2005) 4286–4298.
12. Cohen, G., Honkala, I., Litsyn, S., Lobstein, A.: Covering Codes. Amsterdam, The Netherlands: North-Holland (1997).
13. Courtois, N., Meier, W.: Algebraic attacks on stream ciphers with linear feedback. Advances in Cryptology - Eurocrypt '03 (Lecture Notes in Computer Science, Springer-Verlag) **2656** (2003) 345–359.
14. Dillon, J. F.: Elementary Hadamard Difference Sets. Ph.D. Thesis, University of Maryland (1974).
15. Ding, C., Xiao, G., Shan, W.: The Stability Theory of Stream Ciphers (Lecture notes in Computer Science, Springer-Verlag) **561** (1991).
16. Dobbertin, H.: Construction of bent functions and balanced Boolean functions with high nonlinearity. Fast Software Encryption (Lecture Notes in Computer Science, Springer-Verlag) **1008** (1994) 61–74.
17. Fontaine, C.: On some cosets of the first-order Reed-Muller code with high minimum weight. IEEE Transactions on Information Theory **45** (1999) 1237–1243.
18. Hong, J., Lee, D. H., Yeom, Y., Han, D.: A new class of single cycle t-functions. Fast Software Encryption (Lecture Notes in Computer Science, Springer-Verlag) **3557** (2005) 68–82.
19. Hou, X.-D., Langevin, P.: Results on bent functions. Journal of Combinatorial Theory, Series A **80** (1997) 232–246.
20. Hou, X.-D.: Cubic bent functions. Discrete Mathematics **189** (1998) 149–161.

21. Iwata, T., Kurosawa, K.: Probabilistic higher order differential attack and higher order bent functions. Advances in Cryptology - Asiacrypt '99 (Lecture Notes in Computer Science, Springer-Verlag) **1716** (1999) 62–74.
22. Kasami, T., Tokura, N.: On the weight structure of Reed-Muller codes. IEEE Transactions on Information Theory **16** (1970) 752–759.
23. Kasami, T., Tokura, N., Azumi, S.: On the weight enumeration of weights less than $2.5d$ of Reed-Muller codes. Information and Control **30** (1976) 380–395.
24. Knudsen, L. R., Robshaw, M. J. B.: Non-linear approximations in linear cryptanalysis. Advances in Cryptology - Eurocrypt '96 (Lecture Notes in Computer Science, Springer-Verlang) **1070** (1996) 224-236.
25. Kohavi, Z.: Switching and Finite Automata Theory. McGraw-Hill Book Company (1978).
26. Kumar, P. V., Scholtz, R. A., Welch, L. R.: Generalized bent functions and their properties. Journal of Combinatorial Theory, Series A **40** (1985) 90–107.
27. Kurosawa, K., Iwata, T., Yoshiwara, T.: New covering radius of Reed-Muller codes for t-resilient functions. Selected Areas in Cryptography '01 (Lecture Notes in Computer Science, Springer-Verlang) **2259** (2001) 75–86.
28. Kurosawa, K., Iwata, T., Yoshiwara, T.: New covering radius of Reed-Muller codes for t-resilient functions. IEEE Transactions on Information Theory **50** (2004) 468–475.
29. Lidl, R., Niederreiter, H.: Finite Fields. vol. 20 of Encyclopedia of Mathematics and its Applications, Cambridge University Press 2nd ed. (1996).
30. MacWilliams, F. J., Sloane, N. J. A.: The Theory of Error Correcting Codes. Amsterdam, The Netherlands: North-Holland (1977).
31. Matsui, M.: Linear cryptanalysis method for DES cipher. Advances in Cryptology - Eurocrypt '93 (Lecture Notes in Computer Science, Springer-Verlag) **765** (1993) 386–397.
32. Meier, W., Staffelbach, O.: Nonlinearity criteria for cryptographic functions. Advances in Cryptology - Eurocrypt '89 (Lecture Notes in Computer Science, Springer-Verlag) **434** (1990) 549–562.
33. Menezes, A. J., Van Oorschot, P. C., Vanstone, S. A.: Handbook of Applied Cryptography. CRC Press (1996).
34. Millan, W. L.: Low order approximation of cipher functions. Cryptology: Policy and Algorithms Conference '95 (Lecture Notes in Computes Science, Springer-Verlag) **1029** (1995) 144–155.
35. Millan, W. L.: Analysis and Design of Boolean Functions for Cryptographic Applications. Ph.D. Thesis, Queensland University of Technology (1997).
36. Pieprzyk, J.: Nonlinearity of exponent permutations. Advances in Cryptology - Eurocrypt '89 (Lecture Notes in Computer Science, Springer-Verlag) **434** (1990) 80–92.
37. Preneel, B., Van Leekwijck, W., Van Linden, L., Govaerts, R., Vandevalle, J.: Propagation characteristics of boolean functions. Advances in Cryptology - Eurocrypt '90 (Lecture Notes in Computer Science, Springer-Verlag) **473** (1991) 161–173.
38. Rothaus, O. S.: On bent functions. Journal of Combinatorial Theory, Series A **20** (1976) 300–305.
39. Siegenthaler, T.: Correlation-immunity of nonlinear combining functions for cryptographic applications. IEEE Transactions on Information Theory **30** (1984) 776–780.