Asymptotic Behaviour of Boolean Functions Nonlinearity in Parallel with First-order Reed-Muller Code's Decoding Capability

Stéphanie DIB*

Institut de Mathématiques de Luminy, Marseille, France

Abstract. The aim of this work is to present an accurate distribution of Boolean functions according to the second-order nonlinearity. A lower bound being already proven for almost all Boolean functions, we specify their distribution by proving a concentration point for the nonlinearity of almost all Boolean functions. When considering the case of the (first) nonlinearity, our method meets the already proven result indeed, that is a concentration point as well. But more importantly, throughout this track, we were led to evaluate the error-correcting capability of the first-order Reed-Muller code below and beyond this concentration point. In turn, a decoding threshold is derived above which the probability of error-correcting vanishes in maximum-likelihood decoding.

Index terms - Boolean functions, nonlinearity, Reed-Muller codes, error-correction capability, decoding threshold.

1 Introduction

The nonlinearity $NL(f)$ of a Boolean function f of n variables is its Hamming distance to the set of affine functions, that is

$$NL(f) = \min_{h \text{ affine}} d_H(f, h).$$

It is used to measure the strength of cryptosystems when facing linear attacks. When dealing with low-degree approximation attacks [6, 11], we consider the higher-order nonlinearities. For every $0 \leq r \leq n$, we denote by $NL_r(f)$ the Hamming distance of f to the set of all functions whose algebraic degrees do not exceed r, that is the Reed-Muller code of order r. We have

$$NL_r(f) = \min_{g \in \mathrm{RM}(r,n)} d_H(f, g).$$

The nonlinearity of a Boolean function f of n variables is bounded from above by $2^{n-1} - 2^{n/2-1}$. When n tends to infinity, it was shown by C. Carlet [2], D. Olejàr and M. Stanek [7], that almost all Boolean functions have nonlinearities greater than $2^{n-1} - c \cdot 2^{\frac{n}{2}-1}\sqrt{2n \ln 2}$, where c is a real number greater than 1. Later on, the distribution of Boolean functions between the lower and upper bound was

* This work has been done with the support of the Région Provence-Côte d'Azur

established by F. Rodier [8–10] who proved that the nonlinearity of almost all Boolean functions lies in a small neighbourhood of $2^{n-1} - 2^{\frac{n}{2}-1}\sqrt{2n\log 2}$. This result has been proven as well later by S. Litsyn and A. Shpunt with a different approach [12].
As for the second-order nonlinearity, we recall that it is bounded from above by $2^{n-1} - \sqrt{15} \times 2^{n/2-1} + O(1)$ as proven by C. Carlet and S. Mesnager [4]. Asymptotically, C. Carlet [3] showed that the density of the set of functions satisfying

$$NL_2(f) > 2^{n-1} - c\ \sqrt{\binom{n}{2}\ln 2}\ 2^{\frac{n-1}{2}},$$

tends to 1 when n tends to infinity, if c is a real number greater than 1.

2 First result

Our result [5] shows that $2^{n-1} - \sqrt{\binom{n}{2}\ln 2}\ 2^{\frac{n-1}{2}}$ represents a concentration point of the second-order nonlinearity of Boolean functions.
The approach is basically the second moment method and applies mainly fundamental combinatorics and probability theory, whereas Rodier's approach involves harmonic analysis.
We address the problem as follows. We investigate the subset of Boolean functions whose second-nonlinearities are greater than $\delta = 2^{n-1} - c\ \sqrt{\binom{n}{2}\ln 2}\ 2^{\frac{n-1}{2}}$ where c is a strictly positive real number less than 1, δ being slightly greater than the lower bound given for almost all Boolean functions. By definition, these functions are those for which the Hamming distance to $\mathrm{RM}(2,n)$ exceeds δ. A geometrical viewpoint is the way to track these functions. Therefore, we consider the Hamming closed balls of radius δ centered at all codewords of $\mathrm{RM}(2,n)$, defined by $B_\delta[g] = \{f \in \mathcal{B}_n \mid d_H(f,g) \leq \delta\}$, where g ranges over $\mathrm{RM}(2,n)$. Notice that such functions are strictly those for which the sum of all indicator functions of the above-mentioned Hamming balls equals 0. Hence, over the set $\mathcal{B}_n$ of equiprobable Boolean functions, the problem of determining the probability of the event $\{NL_2(f) > \delta\}$ is equivalent to that of the event $\{\eta = 0\}$, where $\eta = \sum_{g\in\mathrm{RM}(2,n)} 1_{B_\delta[g]}$ is an integer-valued function defined on $\mathcal{B}_n$. This is followed by application of the Chebyshev's inequality [1, chapter IV]

$$P(\eta = 0) \leq P\big(|\eta - E(\eta)| \geq E(\eta)\big) \leq \frac{E(\eta^2) - E(\eta)^2}{E(\eta)^2}.$$

This upper bound allows us to prove a concentration point around
$2^{n-1} - \sqrt{\binom{n}{2}\ln 2}\ 2^{\frac{n-1}{2}}$, when n tends to infinity.
Through this path, we deal with the number of Boolean functions inside a Hamming ball and the intersection between two of them as well. In the (first) nonlinearity case, these estimates yield the following result.

3 Second result

Our second result shows that the concentration point related to the nonlinearity case, that is $2^{n-1} - 2^{\frac{n}{2}-1}\sqrt{2n \log 2}$, is moreover a decoding threshold of the first-order Reed-Muller code when considering maximum-likelihood decoding. Indeed, it is possible to decode correctly almost always errors of weight up to $2^{n-1} - 2^{\frac{n}{2}-1}\sqrt{2n \log 2}$ but the decoding fails almost always when error weight exceeds this value. This bound is unrelated to any algorithm but proves that one can't design algorithm to enhance the error-correction capability of first-order Reed-Muller code further from this threshold.

References

1. N. Alon, Noga; J. Spencer, The probabilistic method. Third edition. With an appendix on the life and work of Paul Erdös. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, 2008.
2. C. Carlet, On cryptographic complexity of Boolean functions, Proceedings of the Sixth Conference on Finite Fields with Applications to Coding Theory, Cryptography and Related Areas. Springer, G.L. Mullen, H. Stichtenoth and H. Tapia-Recillas Eds, pp. 53-69, 2002.
3. C. Carlet, The complexity of Boolean functions from cryptographic viewpoint, preprint.
4. C. Carlet, S. Mesnager, Improving the upper bounds on the covering radii of binary Reed-Muller codes. IEEE Trans. Inform. Theory 53 (2007), no. 1, 162–173.
5. S. Dib, Distribution of Boolean functions according to the second-order nonlinearity, to be published with the proccedings of the International Workshop on the Arithmetic of Finite Fields in the Springer lecture Notes in Computer Science.
6. L. Knudsen, M. Robshaw, Nonlinear approximations in linear cryptanalysis, EUROCRYPT 96, Lecture Notes in Computer Science, 1070, Springer Berlin / Heidelberg, 224-236.
7. D. Olejár et M. Stanek, On cryptographic properties of random Boolean functions, J.UCS 4 (1998), no. 8, 705–717.
8. F. Rodier, Sur la non-linéarité des fonctions booléennes, Acta Arithmetica, vol 115, (2004), 1-22, prétirage de l'IML n° 2002-07, disponible sur ArXiv: math.NT/0306395.
9. F. Rodier, On the nonlinearity of Boolean functions, Proceedings of WCC2003, Workshop on coding and cryptography 2003, sous la direction de D. Augot, P. Charpin, G. Kabatianski, INRIA, 2003, pp. 397 – 405.
10. F. Rodier, Asymptotic nonlinearity of Boolean functions, Designs, Codes and Cryptography, 40:1 2006.
11. T. Shimoyama1, T. Kaneko, Quadratic Relation of S-box and Its Application to the Linear Attack of Full Round DES, CRYPTO '98, Lecture Notes in Computer Science, 1462/, Springer Berlin / Heidelberg (1998), 129-147.
12. S. Litsyn, A. Shpunt, On the distribution of Boolean function nonlinearity. SIAM J. Discrete Math. 23 (2008/2009), no. 1, 79–95.