CLASSIFICATION OF BOOLEAN FUNCTIONS

Valérie Gillot, Philippe Langevin

CLASSIFICATION OF BOOLEAN FUNCTIONS

VALÉRIE GILLOT, PHILIPPE LANGEVIN

Abstract. This note presents a descending method that allows us to classify quotients of Reed-Muller codes of lenghth 128 under the action of the affine general linear group.

1. introduction

Let $\mathbb{F}_2$ be the finite field of order 2. Let m be a positive integer. A mapping from $\mathbb{F}_2^m$ into $\mathbb{F}_2$ is called a Boolean function. Every Boolean function has a unique algebraic reduced representation :

$$f(x_1, x_2, \ldots, x_m) = f(x) = \sum_{S \subseteq \{1,2,\ldots,m\}} a_S X_S, \quad a_S \in \mathbb{F}_2,\ X_S(x) = \prod_{s \in S} x_s.$$

The degree of f is the maximal cardinality of S with $a_S = 1$ in the algebraic form. The valuation of $f \neq 0$, denoted by $\mathrm{val}(f)$, is the minimal cardinality of S for which $a_S = 1$. Conventionnally, $\mathrm{val}(0)$ is ∞. We denote by $B(s,t,m)$ the space of Boolean functions of valuation greater than or equal to s and of degree less than or equal to t. Note that $B(s,t,m) = \{0\}$ whenever $s > t$. The space $B(0,t,m)$ identifies with the Reed-Muller code $RM(t,m)$. The Reed-Muller codes are nested and $B(s,t,m)$ is the representation of the quotient space $RM(t,m)/RM(s-1,m)$. The affine general linear group of $\mathbb{F}_2^m$, denoted by $\mathrm{AGL}(m,2)$, acts naturally over all these spaces. The number of classes of $B(s,t,m)$, denoted by $\mathrm{n}(s,t,m)$, satisfies a nice duality relation :

$$\mathrm{n}(s,t,m) = \mathrm{n}(m-t, m-s, m). \tag{1}$$

X.-D. Hou gives a proof of the above relation in [3]. In the proof of Lemma 1, we propose an alternative demonstration.

For the dimensions that we want to consider, all class numbers are very easy to determine using Burnside's Lemma and the theory of conjugacy classes of $\mathrm{AGL}(m,2)$, see e.g. [4].

In general, such a class number is huge, but, when it is reasonably small, one may consider to determine an orbit representative set that is a list of $\mathrm{n}(s,t,m)$ Boolean functions, of degree less than or equal to t, and pairwise non affine equivalent modulo $RM(s-1,m)$. As an example, the class number $\mathrm{n}(2,6,6)$ is 150357. J. Maiorana in [5] describes a recursive algorithm to find the 150357 equivalence classes.

More generally, the classification information of the space $B(s,t,m)$ plays an important role both in coding theory and cryptography. Indeed, the covering radii of Reed-Muller codes are not generally known. The classification of $B(s,t,m)$

Date: december 2021.

can be used to bound the covering radius of $RM(s-1,m)$ as in the paper [8]. These classifications are also used to study the cryptographic parameters of Boolean functions.

This paper presents a procedure to provide classifications of Boolean functions spaces for $m = 7$. Precisely, we compute orbit representative sets of $B(s,t,7)$, for all parameters $s \leq t \leq 7$ such that $\mathrm{n}(s,t,7)$ is less than 10^6.

Our approach gives *complete* classifications : not only sets of orbit representatives, but also for each representative, a generator set of stabilizer group.

The computed data are available on the project page [2].

2. Boolean functions

A Boolean function f is a member of $B(s,t,m)$ if and only if $s \leq \mathrm{val}(f)$ and $\deg(f) \leq t$. Denoting $\bar{S}$ the complement set of $S \subseteq \{1,2,\ldots,m\}$, the complementary transform $\sum_S X_S \mapsto \sum_S X_{\bar{S}}$ maps $B(s,t,m)$ onto $B(m-t,m-s,m)$, in particular, these spaces have the same dimension. The Reed-muller spaces are nested :

$$\overbrace{RM(-1,m)}^{(0)} \subset RM(0,m) \subset RM(1,m) \subset \cdots \subset RM(m-1,m) \subset \underbrace{RM(m,m)}_{B(m)}.$$

The quotient space $RM(r,m)/RM(r-1,m)$ is the space of forms of degree r, identified with the space $B(r,r,m)$ but we prefer to introduce a notation $H(r,m)$, its dimension is given by the binomial coefficient $\binom{m}{r}$. The dimension of $B(s,t,m)$ is equal to the sum of binomial coefficients $\sum_{w=s}^{t}\binom{m}{w}$. It is easy to see that the weight of a Boolean function is even if and only if its degree is not maximal, consequently the orthogonal of $RM(k,m)$ is $RM(m-k-1,m)$, with respect to the scalar product $\langle f,g\rangle = \sum_{x\in\mathbb{F}_2^m} f(x)g(x)$.

Lemma 1 (duality)**.** *For all s,t such that $s \leq t \leq m$, $B(m-t,m-s,m)$ is a representation of $B(s,t,m)^*$, the dual of $B(s,t,m)$. It means that for any form $\phi \in B(s,t,m)^*$ there exists one and only one $g \in B(m-t,m-s,m)$ such that $\varphi(f) = \langle f,g\rangle$, for all $f \in B(s,t,m)$.*

Proof. Note that the dimension of $B(m-t,m-s,m)$ is precisely the dimension of $B(s,t,m)$. If $0 \neq g \in B(m-t,m-s,m)$ then $g \notin B(s,t,m)^{\perp}$. Indeed, consider a monomial term X_S of maximal degree in the algebraic representation of g. The product $X_{\bar{S}}g$ has degree m whence $X_{\bar{S}}$ is member of $B(s,t,m)$ which is not orthogonal to g. In other words, the space $B(m-t,m-s,m)$ is a representation of $B(s,t,m)^*$. □

The order of $\mathrm{AGL}(m,2)$ is $2^m\prod_{i=0}^{m-1}(2^m-2^i) \approx 0.29\ 2^{m^2+m}$. The affine general linear group acts naturally on the right over Boolean functions. The action of $\mathfrak{s} \in \mathrm{AGL}(m,2)$ on a Boolean function f is $f\circ\mathfrak{s}$. Note that the rank of this action has doubly exponential growth with the parameter m. For $m = 7$, it is already numerically impossible to list the $\approx 2^{74}$ classes.

Two Boolean functions f and g in m variables are equivalent at level r if there exists $\mathfrak{s} \in \mathrm{AGL}(m,2)$ such that $\deg(g + f\circ\mathfrak{s}) \leq r$. We use the notations $f \underset{r}{\sim} g$ for the equivalence at level r, and $\mathrm{STAB}_m^r(f)$, for the stabilizer of f at level r :

$$\text{STAB}_m^r(f) = \{\mathfrak{s} \in \text{AGL}(m,2) \mid f \circ \mathfrak{s} + f \in RM(r,m)\}. \tag{2}$$

The Reed-Muller spaces are invariant under the action of $\text{AGL}(m,2)$. Thus, the affine general linear group acts over the $B(s,t,m)$, the corresponding class number $\mathrm{n}(s,t,m)$ is given by Burnside's formula :

$$|\text{AGL}(m,2)| \times \mathrm{n}(s,t,m) = \sum_{\mathfrak{s}\in\text{AGL}(m,2)} \text{fix}_m^{s,t}(\mathfrak{s}) = \sum_{\mathfrak{s}\in\Gamma} R(\mathfrak{s})\,\text{fix}_m^{s,t}(\mathfrak{s}). \tag{3}$$

where $\text{fix}_m^{s,t}(\mathfrak{s})$ is the number of Boolean functions f of valuation greater than or equal to s and degree less than or equal to t such that $\deg(f\circ\mathfrak{s}+f) < s$ and that is precisely the cardinality of the kernel of the endomorphism of $B(s,t,m)$ defined by $f \mapsto f\circ\mathfrak{s}$. In practice, the sum range Γ is set of representatives of conjugacy classes of $\text{AGL}(m,2)$, and $R(\mathfrak{s})$ the size of the conjugacy class of $\mathfrak{s}$, see book [4] for the finite fields combinatoric details.

Lemma 2 (formula)**.** *For all s,t such that $s \le t \le m$,*

$$\mathrm{n}(s,t,m) = \mathrm{n}(m-t, m-s, m)$$

Proof. Let us recall that the rank of a subgroup of the general linear group over a finite space is the same that the dual group. For $\mathfrak{s} \in \text{AGL}(m,2)$, the adjoint of the automorphism $f \mapsto f\circ\mathfrak{s}$ corresponds to the inverse of $\mathfrak{s}$, because

$$\langle f\circ\mathfrak{s}, g\rangle = \sum_{x\in\mathbb{F}_2^m} f\circ\mathfrak{s}(x)g(x) = \sum_{x\in\mathbb{F}_2^m} f(x)g\circ\mathfrak{s}^{-1}(x) = \langle f, g\circ\mathfrak{s}^{-1}\rangle.$$

The result follows using Burnside's formula by observing

$$\sum_{f,g}(-1)^{\langle f\circ\mathfrak{s}+f,g\rangle} = \sharp B(m-t,m-s,m) \times \text{fix}_m^{s,t}(\mathfrak{s}).$$

□

In this paper, by a classification at level r of degree k in m variables, we want to consider a set of orbit representatives of $B(r+1,k,m)$ under the right action of $\text{AGL}(m,2)$, and for each orbit representative f, the generator system of $\text{STAB}_m^{r+1}(f)$. It is important to note that at level r, we calculate modulo $RM(r,m)$, and we consider polynomials whose valuations are strictly greater than r.

Recall that $\text{AGL}(m,2)$ can be generated by three following transformations : the shift operator $S\colon v \mapsto (v_{m-1},\ldots,v_1,v_m)$, the transvection $T\colon v \mapsto (v_m,\ldots,v_2,v_1+v_2)$ and the translation $U\colon v \mapsto v+(0,\ldots,0,1)$.

In next section, we detail the procedure that we used to build a classification at level $r-1$ from a classification at level r. Starting at level m, there is only one orbit $\{0\}$ stabilized by full group $\text{AGL}(m,2) = \langle S,T,U\rangle$. One can start from this classification at level m to determine the classifications at level $m-1$, level $m-2$ etc.

3. Descending procedure

In order to deduce a classification at level $r-1$ from a classification a level r, we have to consider some "boundary actions" on $H(r,m)$ the space of homogeneous forms of degree r. The stabilizer of f at level r induces an action on homogeneous

polynomials of degree r by mapping $u \in H(r,m)$ to $u \circ \mathfrak{s} + f_{\mathfrak{s}}$ where $f_{\mathfrak{s}}$ is the boundary form $f \circ \mathfrak{s} + f \mod RM(r-1,m)$.

Lemma 3 (boundary)**.** *Let $\mathcal{R}$ be a set of orbit representatives of degree k at level r. For each $f \in \mathcal{R}$, $\mathcal{U}(f)$ denotes a set of orbit representatives of $H(r,m)$ under the boundary action of $\mathrm{STAB}_m^r(f)$. We obtain that $\{f+u \mid f \in \mathcal{R}, u \in \mathcal{U}(f)\}$ is a set of orbit representatives with same degree at level $r-1$.*

Proof. We start by showing the elements of this set are not equivalent at level $r-1$. Indeed, let f' and f be in $\mathcal{R}$, and two forms $u' \in \mathcal{U}(f')$ and $u \in \mathcal{U}(f)$ such that $f+u \underset{r-1}{\sim} f'+u'$. There exists $\mathfrak{s} \in \mathrm{AGL}(m,2)$ such that $f'+u' \equiv (f+u)\circ\mathfrak{s} \mod RM(r-1,m)$. Reducing more, we obtain $f' \equiv f \circ \mathfrak{s} \mod RM(r,m)$; so that f' and f are equivalent at level r, thus $f'=f$. The boundary action of $\mathfrak{s} \in \mathrm{STAB}^r(f)$ sends u to u' and finally $u'=u$. Now, we prove that the set represents all polynomials at level $r-1$. Indeed, let $g \in B(r-1,k,m)$ there exists a pair $(\mathfrak{t},f) \in \mathrm{AGL}(m,2)\times\mathcal{R}$ such that $g\circ\mathfrak{t} \equiv f \mod RM(r,m)$ whence $g\circ\mathfrak{t} \equiv f+v \mod RM(r-1,m)$ where v is a form of degree r, and there is a boundary action $\mathfrak{s} \in \mathrm{STAB}_m^r(f)$ that sends v to some $u \in \mathcal{U}(f)$ whence $g\circ\mathfrak{t}\mathfrak{s} \equiv (f+v)\circ\mathfrak{s} \equiv f+u \mod RM(r-1,m)$. □

Considering the right action of a group G over a set U, we denote by $u \bullet s$ the action of $s \in G$ on $u \in U$, $\mathcal{O}_u$ the orbit of u, S_u the stabilizer of u and o_u the order of S_u.

Lemma 4 (class formula)**.** *If G is a finite group acting on a finite set U then the size of the orbit of an element $u \in U$ is equal to $|G|/o_u$ where o_u is the order of the stabilizer S_u of u.*

Proof. There is a bijection from G/S_u onto $\mathcal{O}_u$ the orbit of u. □

Lemma 5 (Schreier)**.** *Let L be a set of generators of a finite group G right acting on a finite set U. Let $\mathcal{O}_u$ be the orbit of some element $u \in U$. If $R\colon \mathcal{O}_u \to G$ is a map such that $u \bullet R(x) = x$ for all $x \in \mathcal{O}_u$ then $\{R(x)\lambda R(x\bullet\lambda)^{-1} \mid \lambda \in L, x \in \mathcal{O}_u\}$ spans the stabilizer S_u of u.*

Proof. See [7]. □

Knowing the value o_u, one can build a generator set of its stablizer S_u applying Schreier's Lemma. We implement this idea in the algorithm `generatorSet` where $*$ denotes the law group and $\bullet$ denotes the action.

Now, we describe our descending procedure based on Lemma 3 and Lemma 5 to construct a set of orbit representatives at level $r-1$ from level r. In view of dimension of forms space $H(r,m)$ and for save memory space, we proceed in two phases :

(1) For each representative f at level r, we use a classical algorithm to enumerate an orbit representatives set of $H(r,m)$ under the action of $\mathrm{STAB}_m^r(f)$. For each representative u, we obtain the orbit $\mathcal{O}_u$, and by Lemma 4, the order o_u of $\mathrm{STAB}_m^{r-1}(f+u)$ is equal to $\sharp\mathrm{STAB}_m^r(f)/\sharp\mathcal{O}_u$.
(2) For each representative f at level r, let L be a generator set of $\mathrm{STAB}_m^r(f)$. For each pair (u,o_u), obtained in (1), we apply `generatorSet(`u`, `L`, `o_u`)` to construct a set of generators of $\mathrm{STAB}_m^{r-1}(f+u)$.

LISTING 1. Construction of a generator set of S_u.

```
1  Algorithm generatorSet( u , L, o_u )
2  {   // return a generator set of the stabilizer of u
3      // under the action of the group generated by L
4      // knowing its order o_u
5      S ←∅
6      push( u )
7      R [ u ] ← id
8      Y ←{ u }
9      while ( order( <S> ) < o_u) {
10         pop( x )
11         for λ ∈ L {
12             y ← x • λ
13             if y ∉ Y {
14                 push(y)
15                 R[ y ] ←R[ x ] ∗ λ
16                 Y← Y ∪ {y}
17             } else {
18               s ← R[x] ∗ λ ∗ inverse ( R[ y ] )
19               if ( s not in <S> )
20                  S ←S ∪ { s }
21             }
22         }
23     return S;
24 }
```

TABLE 1. Class numbers $\mathrm{n}(s,t,7)$.

$s\backslash t$	1	2	3	4	5	6	7
0	3	12	3486	$10^{13.5}$	$10^{19.8}$	$10^{21.9}$	$10^{22.2}$
1	2	8	1890	$10^{13.1}$	$10^{19.5}$	$10^{21.6}$	$10^{21.9}$
2		4	179	$10^{11.0}$	$10^{17.3}$	$10^{19.5}$	$10^{19.8}$
3			12	68443	$10^{11.0}$	$10^{13.1}$	$10^{13.5}$
4				12	179	1890	3486
5					4	8	12
6						2	3
7							2

4. APPLICATION

An alternative way to build a list of orbit representatives consists to use invariants. Success for invariant based approach is not guaranteed for two reasons : small orbits are hidden and difficult to detect, and the invariants used may not be discriminating enough ! Moreover, invariant approach does not give orbit sizes and even less the generator set of stabilizers. The invariant approach in [6] failed to find a list of representatives of $B(2,4,7)$. In that case, the number of orbits is

$\mathrm{n}(2,4,7) = 68433$ and using invariants, the authors got 68095 orbits missing 338 orbits.

Our implementation in C language of the descending procedure, without any parallelization, builds the full classification of $B(2,6,6)$ in 15 secondes. It find the classification of $B(2,4,7)$ in three days and require about 50GB of memory. The values of $\mathrm{n}(s,t,7)$ for $0 \leq s \leq t \leq 7$ are listed in Table . For all parameters $0 \leq s \leq t \leq 7$ such that $\mathrm{n}(s,t,7) < 10^6$, the descending procedure classifies $B(s,t,7)$, it computes for each orbit, a representative and also a generator sets of the corresponding stabilizer. All the numerical data are available in project page.

On the side of coding theory, writting this note, we learned that the covering radius of $RM(3,7)$ has just been determined in the paper [1]. At the same time, we used the classification of $B(4,7,7)$ to obtain the value 20 for the covering radius.

5. conclusion

Our descending procedure successfully classifies Boolean functions in 7 variables. The dimension of form spaces is an obstruction to apply this approach in dimension 8. Perhaps, it is possible to mix the descending procedure and the recursive approach of Maiorana to obtain classification results for higher dimension.

References

[1] J. Gao, H. Kan, Y. Li, and Q. Wang. The covering radius of the third-order reed-muller codes $\mathrm{rm}(3,7)$ is 20. *submitted to IEEE IT*, 2023.

[2] Valérie Gillot and Philippe Langevin. Classification of $b(s,t,7)$. `http://langevin.univ-tln.fr/project/agl7/aglclass.html`, 2022.

[3] Xiang-Dong Hou. $\mathrm{AGL}(m,2)$ acting on $R(r,m)/R(s,m)$. *J. Algebra*, 171(3):921–938, 1995.

[4] Xiang-Dong Hou. *Lectures on finite fields*, volume 190 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2018.

[5] James A. Maiorana. A classification of the cosets of the Reed-Muller code $R(1,6)$. *Math. Comp.*, 57(195):403–414, 1991.

[6] Meng Qingshu, Zhang Huanguo, Cui Jingsong, and Yang Min. Almost enumeration of eight-variable bent functions. *iacr preprint*, 2005.

[7] Ákos Seress. *Permutation group algorithms*, volume 152 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2003.

[8] Qichun Wang. The covering radius of the Reed-Muller code $RM(2,7)$ is 40. *Discrete Math.*, 342(12):111625, 7, 2019.

Imath, université de Toulon