Journal of Integer Sequences, Vol. 10 (2007),
Article 07.3.4

Mahler's Expansion and Boolean Functions

Jean-Francis Michon[1]
LITIS - Université de Rouen
Avenue de l'Université - BP 8
76801 Saint-Étienne-du-Rouvray Cedex
France
jean-francis.michon@univ-rouen.fr

Pierre Valarcher
LACL - Université Paris 12 Val de Marne
Faculté des Sciences
61, avenue du Général de Gaulle
94010 Créteil Cedex
France
valarcher@univ-paris12.fr

Jean-Baptiste Yunès
LIAFA - Université Denis Diderot Paris 7
175, rue du chevaleret
75013 Paris
France
Jean-Baptiste.Yunes@liafa.jussieu.fr

Abstract

The substitution of X by X^2 in binomial polynomials generates sequences of integers by Mahler's expansion. We give some properties of these integers and a combinatorial interpretation with covers by projection. We also give applications to the classification of boolean functions. This sequence arose from our previous research on classification and complexity of Binary Decision Diagrams (BDD) associated with boolean functions.

[1]All correspondence should be directed to this author.

1 Mahler's expansion

We recall some standard facts about binomial polynomials and Mahler's expansion (see [2, 3, 5].

A binomial polynomial $B_j(X) = \binom{X}{j} \in \mathbb{Q}[X]$, for any integer $j \geq 1$, is defined by:

$$B_j(X) = \frac{X(X-1)\cdots(X-j+1)}{j!}$$

and $B_0(X) = \binom{X}{0} = 1$ by convention. For example: $\binom{X}{1} = X$, $\binom{X}{2} = \frac{X(X-1)}{2}$.

The degree of B_j is j, so they form a **basis** of $\mathbb{Q}[X]$. The expansion of a polynomial in this basis is called its **Mahler expansion**, also known as the Newton interpolation formula.

From the definition, the j roots of B_j are $0, \ldots, j-1$. This can be interpreted as an extension of the definition of binomial coefficients: for $n, j \in \mathbb{N}$, $\binom{n}{j} = 0$ if $n < j$.

The Pascal triangle equality is

$$\binom{X+1}{j} = \binom{X}{j} + \binom{X}{j-1}$$

for $j > 0$. This equality says that, in this basis, the endomorphism $P(X) \to P(X+1)$ of $\mathbb{Q}[X]$ has a Jordan form.

Let f any function from $\mathbb{Q}_p \to \mathbb{Q}_p$, where $\mathbb{Q}_p$ is the field of p-adic numbers, using the difference operators:

$$\begin{aligned}
\Delta f &= f(X+1) - f(X) \\
\Delta^2 f &= f(X+2) - 2f(X+1) + f(X) \\
&\vdots \\
\Delta^j f &= \sum_{r=0}^{j} (-1)^r \binom{j}{r} f(X+j-r).
\end{aligned}$$

Then the Mahler expansion of f is

$$\sum_{j=0}^{\infty} (\Delta^j f)(0) \binom{X}{j} \tag{1}$$

Mahler's theorem says that, for any prime p, this expansion converges uniformly towards f if f is any continuous mapping $f : \mathbb{Z}_p \to \mathbb{Q}_p$.

2 Squaring variable operator

Consider the $\mathbb{Q}$-linear endomorphism of $\mathbb{Q}[X]$ defined by:

$$f(X) \mapsto f(X^2) \tag{2}$$

This endomorphism is clearly injective because $f(X^2) = g(X^2)$ implies that $f - g$ has infinitely many roots: the squares of $\mathbb{Q}$. It is also an algebra endomorphism because $(fg)(X^2) = f(X^2)g(X^2)$ and constant 1 is invariant.

We study the effect of this operator on the basis B_j.

Definition 1. *We define $a_{k,m} \in \mathbb{Q}$ for all $k, m \in \mathbb{N}$ as the coefficients of the Mahler expansion of $\binom{X^2}{k}$:*

$$B_k(X^2) = \binom{X^2}{k} = \sum_{m=0}^{\infty} a_{k,m} B_m \tag{3}$$

This double sequence is the sequence A100344 in Sloane's Online Encyclopedia [6].

2.1 General properties of the $a_{k,m}$

For fixed k, all the $a_{k,m}$ are 0, except a finite number of them.

From this definition we compute the first values of $a_{k,m}$:

$$a_{0,0} = 1, \quad a_{0,r} = a_{r,0} = 0 \text{ for } r > 0$$

We give in the last section the table for the first values of $a_{k,m}$. The binomial polynomials $B_j(X)$ are a $\mathbb{Z}$-basis for the $\mathbb{Z}$-module of integer polynomials, therefore trivially $a_{k,m} \in \mathbb{Z}$, but we shall now prove that $a_{k,m} \in \mathbb{N}$.

From (3), substituting X with all integral values in $\mathbb{N}$, we get an infinite linear system. The study of this system gives many important properties of the $a_{k,m}$.

Proposition 1. *$a_{k,m} = 0$ if $m > 2k$ or $m < \sqrt{k}$.*

Proof. To establish the first inequality consider that the degree of the left-hand side of (3) is $2k$.

The second inequality is obviously true for $k = 0$. Suppose $k > 0$, if $n \in \mathbb{N}$ and $n < \sqrt{k}$ then $n^2 < k$ and n is a root of $\binom{X^2}{k}$. Replacing X by n in (3) with $n = 0, \dots, m$ we get

$$\begin{aligned} 0 &= a_{k,0} \\ 0 &= a_{k,0} + a_{k,1} \\ 0 &= a_{k,0} + 2a_{k,1} + a_{k,2} \\ &\vdots \\ 0 &= a_{k,0} + \binom{m}{1} a_{k,1} + \binom{m}{2} a_{k,2} + \dots + a_{k,m} \end{aligned}$$

and so $a_{k,0} = a_{k,1} = \dots = a_{k,m} = 0$. $\square$

Proposition 2 (First recursive formula)**.** *For all $k, n \in \mathbb{N}$*

$$a_{k,n} = \binom{n^2}{k} - \sum_{m=0}^{n-1} a_{k,m} \binom{n}{m} \tag{4}$$

Proof. We suppose first $0 \leq n \leq 2k$. We use (3) and write $\binom{X^2}{k} = \sum_{m=0}^{2k} a_{k,m} B_m$. Make $X = n$ and use the property that $B_m(n) = 0$ if $m > n$, and $B_n(n) = 1$.

If $n > 2k$ we must show that the right-hand side of (4) is 0. But $\sum_{m=0}^{n-1} a_{k,m} B_m = \sum_{m=0}^{2k} a_{k,m} B_m$ by Proposition 1 and this sum is 0 by definition of the $a_{k,m}$. $\square$

A consequence of this Proposition is that $a_{k,m} \in \mathbb{Z}$.

Proposition 3. *For all integers k, m we have*

$$a_{k,m} = \sum_{i=0}^{m} (-1)^{m-i} \binom{m}{i} \binom{i^2}{k} \tag{5}$$

Proof. This is a just a translation of Mahler's coefficient computation (1). $\square$

Proposition 4. *(a)* $a_{k,2k} = \frac{(2k)!}{k!}$

(b) $a_{k,m} = \binom{m^2}{k}$ *if* $k > (m-1)^2$

Proof. The first identity (a) is easily obtained by comparing the leading coefficients of the polynomials of left-hand side and right-hand side of (3) which are $\frac{1}{k!}$ and $a_{k,2k} \frac{1}{(2k)!}$ respectively.

To prove the second equality (b), use (4) and Proposition 1, which shows that all the terms in the sum are 0. $\square$

It is useful for computing to formally generalize the definition when m is a negative integer. We shall set $a_{k,m} = 0$ if $k \in \mathbb{N}$ and $m < 0$.

We now prove a more difficult identity:

Theorem 1 (Second recursive formula)**.** *For $k \geq 1$*

$$a_{k,m} = \frac{1}{k}[(m^2 - k + 1)a_{k-1,m} + m(2m-1)a_{k-1,m-1} + m(m-1)a_{k-1,m-2}].$$

Proof. Consider the endomorphism of $\mathbb{Q}[X]$ defined by $f(X) \to Xf(X)$ (multiplication by X). We study its effect on the B_m basis. Clearly, for all $m \geq 0$

$$XB_m = (X - m + m)B_m = (m+1)B_{m+1} + mB_m$$

We consider now the endomorphism $f(X) \to X^2 f(X)$ (multiplication by X^2). Its effect on the binomial basis is, by iteration of the preceding formula

$$\begin{aligned} X^2 B_m &= (m+1)[(m+2)B_{m+2} + (m+1)B_{m+1}] + m(m+1)B_{m+1} + m^2 B_m \\ &= (m+1)(m+2)B_{m+2} + (m+1)(2m+1)B_{m+1} + m^2 B_m \end{aligned}$$

We start from ($k \geq 1$):

$$\binom{X^2}{k} = \binom{X^2}{k-1} \frac{X^2 - k + 1}{k}$$

and expand the right-hand side

$$\frac{X^2-k+1}{k}\sum_{m=0}^{2k-2} a_{k-1,m}B_m$$

$$=\frac{1}{k}\sum_{m=0}^{2k-2} a_{k-1,m}[(m+1)(m+2)B_{m+2}+(m+1)(2m+1)B_{m+1}+(m^2-k+1)B_m]$$

Grouping together the coefficients of B_m we get the formula. ☐

We applied this formula to construct the table in Section 6. We started from the first column et derived all others.

Corollary 1. $0 \leq a_{k,m} \leq \binom{m^2}{k}$

Proof. From Theorem 1 if $m^2-k+1<0$ or if $m=0$ then $a_{k,m}=0$ or 1, in all other cases the coefficient used in Theorem 1 are ≥ 0. The higher bound is an immediate consequence of the positivity and of the recurrence formula. ☐

Corollary 2. $a_{k,2k-1} = a_{k,2k} \cdot \frac{2k-1}{2} = \frac{(2k)!}{k!}\frac{2k-1}{2}$

Proof. Easy consequence. ☐

Corollary 3. *Fix m, then the sequence $a_{k,m}$ is increasing with k for $0 \leq k \leq \frac{m^2}{2}$.*

Proof. By Theorem 1, for $k>0$:

$$a_{k,m} \geq \frac{m^2-k+1}{k}a_{k-1,m}$$

and $\frac{m^2-k+1}{k} = \frac{m^2}{k} - 1 + \frac{1}{k} \geq 1 + \frac{1}{k} > 1$. ☐

Questions: Fix m or k, prove the $a_{k,m}$ are increasing then decreasing and find good bound for them. Are there other simple expressions for the $a_{k,m}$?

3 Covering of a finite set by projection

Let $M=[1..m]$ the set of integers from 1 to m, and a fixed integer k. We look for families F of k distinct pairs $F=\{(a_1,b_1),\ldots,(a_k,b_k)\}\subset M^2$. If $\bigcup_{i=1}^{k}\{a_i,b_i\}=M$ we say that F is a **covering of M by projection**.

This definition easily generalizes to any exponent r of M; in this way, we get families of k distinct r-uples covering of M^r by projection. All the results of this article could have been written in this perspective.

This allows a straightforward combinatorial interpretation of the $a_{k,m}$.

Theorem 2. *The number of parts $F \subset M^2$ of k distinct pairs covering M by projection is $a_{k,m}$.*

Proof. Let $\mathcal{X}$ any finite set with X elements. The number of subsets of $\mathcal{X}^2$ having k elements is $\binom{X^2}{k}$. Each of these subsets is a covering of some subset $M \subset \mathcal{X}$ with m elements by projection and m may take values between 0 and X^2. This enumeration gives each term of the sum in the right-hand side of (3). The coefficients $a_{k,m}$ are uniquely determined by (3) because the binomial polynomials form a basis of the polynomial ring $\mathbb{Q}[X]$. ☐

4 Profiles of boolean functions in n variables

The set $\mathcal{B}_n$ of boolean functions in n variables is the set of all functions

$$f : \{0,1\}^n \to \{0,1\}.$$

It is in bijective correspondence with the set of parts of $\{0,1\}^n$. For $n = 0$, $\mathcal{B}_0$ is the set of the two constant boolean functions 0 and 1. The number of elements of $\mathcal{B}_n$ is 2^{2^n} for all $n \in \mathbb{N}$.

Definition 2. *Let $\mathcal{F} = \{f_1, \dots, f_r\} \subset \mathcal{B}_n$ any finite and non-empty family of distinct boolean functions in n variables. We associate with $\mathcal{F}$ a sequence of $n+1$ positive integers*

$$p(\mathcal{F}) = (p_0(\mathcal{F}), \dots, p_n(\mathcal{F})),$$

*where $p_i(\mathcal{F})$ is the number of distinct boolean functions in $n-i$ variables obtained from $\mathcal{F}$ by substituting all possible boolean values to the **first** i boolean variables $x_1, \dots, x_i$. For $i = 0$ we set $p_0(\mathcal{F}) = r$.*

*We call $p(\mathcal{F})$ the **profile** of the family $\mathcal{F}$ or the profile of f if $\mathcal{F}$ is reduced to one boolean function f.*

> Example with $r = 1$, $\mathcal{F} = \{f(x_1, x_2, x_3) = x_2\}$.
>
> We have $f(0, x_2, x_3) = f(1, x_2, x_3) = x_2$, so $p_1(f) = 1$. If we give all boolean values to x_1 and x_2 (4 possible pairs of values), in all cases we get the 0 (resp. 1) constant function if $x_2 = 0$ (resp. $x_2 = 1$), so we have $p_2(f) = 2$. When we give any boolean values to the three variables we get the constants 0 or 1. Finally $p(f) = (1, 1, 2, 2)$.

The profile is a very interesting "classifier" which is connected to complexity questions. It is related to the Binary Decision Diagram theory (a BDD is a boolean graph canonically associated with any boolean function). A way to define complexity of $f \in \mathcal{B}_n$ is to consider its profile $p(f) = (1, p_1, \dots, p_n)$ and to define its complexity as

$$c(f) = p_0 + \cdots + p_n$$

This complexity measures the number of different "subfunctions" inside f generated by our sequential affectations of values to the variables. In BDD theory it is the number of vertices of the canonical boolean graph associated with f. We refer the reader to our paper [4] for the details and related results.

The important thing about the profile and the complexity, is that they are not invariant by permutations of variables in general. This can be easily verified on our example: if $f(x_1, x_2, x_3) = x1$ then $p(f) = (1, 2, 2, 2)$ and if $f(x_1, x_2, x_3) = x_3$ then $p(f) = (1, 1, 1, 2)$.

Now we can state our main result:

Theorem 3. *The number of families of boolean functions in $n \geq 1$ variables whose profile is $(p_0, \dots, p_n)$ is the product*

$$a_{p_0,p_1} a_{p_1,p_2} \cdots a_{p_{n-1},p_n}$$

with $p_n = 1$ or 2. For $n = 0$ the number is $a_{p_0,1}$.

Proof. The last profile value p_n of any boolean function in n variables is always 1 or 2 because there are only 2 boolean constant functions namely 0 and 1.

We proceed by recurrence. For $n = 0$ the Theorem is true by simple inspection.

The number of families of p_1 distinct boolean functions in the variables $x_2, \ldots, x_n$ whose profile is $(p_1, \ldots, p_n)$ is $a_{p_1,p_2} \cdots a_{p_{n-1},p_n}$ by the recurrence hypothesis. Let $\mathcal{F}' = \{f'_1, \ldots, f'_{p_1}\}$ such a family. Then we can construct an unique boolean function $f \in \mathcal{B}_n$ from each pair $(u, v) \in \mathcal{F}'^2$ by using the well known Boole identity

$$f(x_1, \ldots, x_n) = (1 - x_1)u \oplus x_1 v$$

and we must choose p_0 distinct pairs (u, v) in $\mathcal{F}'^2$. In this manner we can construct a family of p_0 distinct boolean functions in n variables whose profile is $(p_0, \ldots, p_n)$. The number of such families constructed from $\mathcal{F}'$ coincides with the combinatorial definition of the a_{p_0,p_1} as a covering of $\mathcal{F}'$ by projection of p_0 elements.

We conclude that for each $\mathcal{F}'$ we can construct a_{p_0,p_1} families with profile $(p_0, ..., p_n)$,and the Theorem is proved. $\square$

We can specialize the last formula with $p_0 = 1$. We get immediately

Remark 1. *The set of $f \in \mathcal{B}_n$ with profile $(1, p_1, \ldots, p_{n-1}, 2)$ has*

$$a_{1,p_1} \cdots a_{p_{n-1},2}$$

elements.

We list all possible profiles for $n \leq 4$ in lexicographical order and give the number of boolean functions with each profile.

$n=0$		$n=1$		$n=2$		$n=3$		$n=4$	
1	2	1, 1	2	1, 1, 1	2	1, 1, 1, 1	2	1, 1, 1, 1, 1	2
		1, 2	2	1, 1, 2	2	1, 1, 1, 2	2	1, 1, 1, 1, 2	2
				1, 2, 2	12	1, 1, 2, 2	12	1, 1, 1, 2, 2	12
						1, 2, 2, 2	72	1, 1, 2, 2, 2	72
						1, 2, 3, 2	144	1, 1, 2, 3, 2	144
						1, 2, 4, 2	24	1, 1, 2, 4, 2	24
								1, 2, 2, 2, 2	432
								1, 2, 2, 3, 2	864
								1, 2, 2, 4, 2	144
								1, 2, 3, 3, 2	864
								1, 2, 3, 4, 2	10368
								1, 2, 4, 2, 2	8928
								1, 2, 4, 2, 2	144
								1, 2, 4, 3, 2	11808
								1, 2, 4, 4, 2	31728

For a given n, we ignore which profile gives the largest number of boolean functions. This question is connected to interesting works on "Shannon effect" (for short: random functions have almost surely a maximal complexity) [1]. Our computations can be seen as a enumerative and effective approach to this problem. We recall that Shannon's theorem is relative to "circuit" complexity, and for this complexity "almost" nothing effective is known about functions achieving maximal complexity.

5 Conclusion

The $a_{k,m}$ numbers where first introduced in a combinatorial way in our article [4]. Theorem 1 was also proved in a combinatorial way. We were unsuccessful in the search of a generating series for these numbers and realized after a while that the Mahler expansion of the $B_k(X^2)$ is an answer. This permits a whole algebraic reinterpretation of the formulas of our article [4] and the enlargement of the scope to all boolean functions, giving the Theorem 3. Moreover, all the results of this paper can be generalized using other exponent than 2, and give other interpretations with non boolean functions and m-ary trees instead of binary trees. We choose to keep our scope restricted to the squaring and to the boolean functions formulas for simplicity. We think that the interested reader will have no great difficulties in constructing more general formulas if needed.

6 Table of the $a_{k,m}$

	k 0	1	2	3	4	5	6	7	8	9	10
m 0	1	0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0	0
2	0	2	6	4	1	0	0	0	0	0	0
3	0	0	18	72	123	126	84	36	9	1	0
4	0	0	12	248	1322	3864	7672	11296	12834	11436	8008
5	0	0	0	300	4800	32550	137900	423860	1017315	1985785	3228720
6	0	0	0	120	7800	121212	1003632	5634360	23963760	82057010	234694416
7	0	0	0	0	5880	235200	3791032	37162384	261418626	1437954784	6506878224
8	0	0	0	0	1680	248640	8280272	141626144	1605962556	13627345424	92665376496
9	0	0	0	0	0	136080	10886400	336616560	6156764640	79330914540	790034244120
10	0	0	0	0	0	30240	8517600	516327840	15590248560	305402753240	4409098539560
11	0	0	0	0	0	0	3659040	512265600	26837228880	812355376800	17025823879944
12	0	0	0	0	0	0	665280	318003840	31638388320	1529756532480	47104037930928
13	0	0	0	0	0	0	0	112432320	25184839680	2058204788640	95321107801920
14	0	0	0	0	0	0	0	17297280	12955662720	1968191184960	142446885060480
15	0	0	0	0	0	0	0	0	3891888000	1307674368000	157084383456000
16	0	0	0	0	0	0	0	0	518918400	574269696000	126281698583040
17	0	0	0	0	0	0	0	0	0	149967417600	71984360448000
18	0	0	0	0	0	0	0	0	0	17643225600	27576361612800
19	0	0	0	0	0	0	0	0	0	0	6369204441600
20	0	0	0	0	0	0	0	0	0	0	670442572800

References

[1] C. Gröpl, Binary decision diagrams for random boolean functions, PhD thesis, Humboldt-Universität zu Berlin, 1999.

[2] K. Mahler, An interpolation series for continuous functions of a p-adic variable, *J. Reine Angew. Math.* **199** (1958), 23–34. Correction, **208** (1961), pp. 70–72.

[3] K. Mahler, *p-adic Numbers and Their Functions*, Cambridge Tracts in Mathematics, Cambridge University Press (second edition), 1981.

[4] J. F. Michon, P. Valarcher, J.-B. Yunès, On maximal QROBDD's of boolean functions, *RAIRO Theor. Infor. Appl.*, **39** (2005), 677-686.

[5] A. Robert, *A Course in p-adic Analysis*, GTM **198**, Springer, 2000.

[6] N. J. A. Sloane, The On-Line Encyclopedia of Integer Sequences.

2000 *Mathematics Subject Classification*: Primary 05A10; Secondary 94C10.
Keywords: Mahler's expansion, binomial polynomials, boolean functions, BDD, enumeration, complexity.

(Concerned with sequence A100344.)

Received July 18 2005; revised version received March 28 2007. Published in *Journal of Integer Sequences*, March 28 2007.

Return to Journal of Integer Sequences home page.