International J.Math. Combin. Vol.3(2022), 61-68

Generating Functions of Group Codes

Nacer Ghadbane

Laboratory of Pure and Applied Mathematics
Department of Mathematics, University of M'sila, Algeria

E-mail: nacer.ghadbane@univ-msila.dz

Abstract: Let Σ^* is the free monoid over a finite alphabet Σ and H a subgroup of a given group G.A *group code* X is the minimal generator of X^* with $X^* = \Psi^{-1}(H)$, where Ψ is a morphism from the free monoid Σ^* to the group G. *A generating function* is just a different way of writing a sequence. Generating functions transform problems about sequence into problems about functions. In this paper, we will give a several formulas for the generating functions of $X$and X^*.

Key Words: Words and languages, free monoid, morphism of monoids, generating function.

AMS(2010): 94A60, 68Q42, 68Q70, 20M05.

§1. Introduction

Codes are an essential tool in information theory, and the theory of variable length codes is firmly related to combinatorics on words. The object of the theory is to study factorisation of words into sequences of words taken from a given set X. In a free monoid X^* generated by a code X there does not exist two distinct factorisations in X for any word. It is not always easy to verify a given set of words is a code. Some examples of the variable length codes are the Huffman coding, Limpel-Zev-Welch code and Arithmetic coding. The thoery of variable length codes takes its origin in the framework of the theory of information, since Shannon's early works in the 1950's. An algebraic theory of codes was subsequently initiated by M. P. Schutzenberger (see [17]). Variable-length codes occur frequently in the domain of data compression. Statistical data compression methods employ variable-length codes, with the short codes assigned to symbols or groups of symbols that appear more often in the data (have a higher probability of occurrence).

In this paper, we are interested in a particular type of variable lengths codes, called group codes, more precisely we describe in terms of different parameters the generating functions of the group codes and their stars.

The remainder of this paper is organized as follows. In Section 2, we introduce the notations for the rest of the paper and give basic definition of terms that will be helpful as we proceed. In Section 3, we show several formulas for the generating functions of the group codes and its

[1]Received August 14, 2022, Accepted September 16, 2022.

stars. Finally, we draw our conclusions in Section 4.

§2. **Preliminaries**

There is an extremely powerful tool in discrete mathematics used to manipulate sequences called generating function. A generating function is just a different way of writing a sequence of numbers. Generating functions transform problems about sequence into problems about functions. This is great because we've got piles of mathematical machinery for manipulating functions. Let $(g_n)_{n\geq 0}$ be a sequence of numbers. The generating function associated to this sequence is the series $G(x) = \sum\limits_{n\geq 0} g_n x^n$. The correspondence between a sequence and its generating function with a double-sided arrow as follows:

$$\langle g_0, g_1, g_2, g_3, \cdots \rangle \longleftrightarrow G(x) = g_0 + g_1 x + g_2 x^2 + g_3 x^3 \cdots .$$

The magic of generating functions is that we can carry out all sorts of manipulations on sequences by performing mathematical operations on their associated generating functions. Let's experiment with various operations and characterize their effects in terms of sequences. Notice that,

1. If $\langle g_0, g_1, g_2, g_3, \cdots \rangle \longleftrightarrow G(x)$ and $a \in \mathbb{R}$ then,

$$\langle g_0, g_1, g_2, g_3, \cdots \rangle \longleftrightarrow G(x).$$

2. If $\langle g_0, g_1, g_2, g_3, \cdots \rangle \longleftrightarrow G(x)$ then

$$\left\langle \overbrace{0, 0, \cdots, 0}^{k\ zeroes}, g_0, g_1, g_2, g_3, \cdots, \right\rangle \longleftrightarrow x^k G(x).$$

3. If $\langle g_0, g_1, g_2, g_3, \cdots, \rangle \longleftrightarrow G(x)$ then $\langle g_1, 2g_2, 3g_3, \cdots, \rangle \longleftrightarrow G'(x)$.

4. If $\langle g_0, g_1, g_2, g_3, \cdots \rangle \longleftrightarrow G(x)$ and $\langle f_0, f_1, f_2, f_3, \cdots \rangle \longleftrightarrow F(x)$ then

$$\langle g_0 + f_0, g_1 + f_1, g_2 + f_2, g_3 + f_3, \cdots \rangle \longleftrightarrow G(x) + F(x).$$

5. If $\langle g_0, g_1, g_2, g_3, \cdots \rangle \longleftrightarrow G(x)$ and $\langle k_0, k_1, k_2, k_3, \cdots \rangle \longleftrightarrow K(x)$ then

$$\langle m_0, m_1, m_2, m_3, \cdots \rangle \longleftrightarrow G(x) K(x),$$

where

$$m_n = g_0 k_n + g_1 k_{n-1} + g_2 k_{n-2} + \cdots + g_n k_0.$$

Now, le us recall the power series expansion of $(1+x)^{\alpha}$, valid for $\alpha \in \mathbb{R}$,

$$(1+x)^{\alpha} = 1 + \alpha x + \cdots + \begin{pmatrix} \alpha \\ k \end{pmatrix} + \cdots,$$

where, by convention, $\begin{pmatrix} \alpha \\ k \end{pmatrix} = \dfrac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!}$.

A semigroup is a pair $(S, \circ)$, where S is a set and $\circ$ an associative binary operation on S. If the set S contains an element 1_S such that $1_S \circ s = s \circ 1_S = s$ for all $s \in S$ we call $(S, \circ, 1_S)$ a monoid and refer to the element 1_S as the one ore the identity.

A semigroup morphism from a semigroup $(S, \circ)$ into a semigroup $(T, \triangle)$ is a mapping $h : S \longrightarrow T$ such that $h(u \circ v) = h(u) \triangle h(u)$.

Let X and Y be two subsets of a semigroup $(S, \circ)$. The product of X and Y is the set $X \circ Y = \{x \circ y : x \in X, y \in Y\}$. We denote by X^+ the subsemigroup generated by X, that is $X^+ = \{x_1 \circ \cdots \circ x_n : n \geq 1, x_1 \in X\}$. If S is a monoid, we also define $X^* = X^+ \cup \{1_S\}$ which is the submonoid of S generated by X.

Let A denote a finite set of symbols. The elements of A are called letters and the set A is called an alphabet. A finite word over A is a finite sequence of letters $u = (a_1, a_2, \cdots, a_n)$ of elements of A denoted by the concatenation $w = a_1 a_2 \cdots a_n$. The integer $n = |w|$ is the length of the word w. For example, the finite sequences 00110 and 110 are two words over the binary alphabet $\{0, 1\}$ with $|00110| = 5$ and $|110| = 3$. The empty sequence () of length 0 is called the empty word and is denoted by ϵ. The set A^* of all words over A equipped with the operation of concatenation has a structure of a monoid with the empty word ϵ as a neutral element, called the free monoid on A. We denote by $A^+ = A^* - \{\epsilon\}$ the free semigroup over A.

For example, $\{0, 1, 2\}^* = \{\epsilon, 0, 1, 2, 00, 01, 02, 11, 12, 20, 21, \cdots\}$. If a is a letter of the alphabet A, for any word $w = a_1 a_2 \cdots a_k$ of A^*, we denote by $|u|_a = Card\,\{i = 1, 2, \cdots, k : a_i = a\}$, the number of the occurrences of a in the word u. For example, we have $|00110|_0 = 3$ and $|00110|_1 = 2$ [8].

For $X \subset A^*$, we define $X^0 = \{\epsilon\}, X^{n+1} = X^n X$ $(n \geq 0)$ and $X^* = \bigcup_{n \geq 0} X^n$. Note that, any submonoid M of A^* has a unique minimal generating set $(M - \epsilon) - (M - \epsilon)^2$.

Given two words $u, w \in A^*$, we say that u is factor (prefix, suffix) of w if and only if we have $w \in A^* u A^*$ $(w \in uA^*, w \in A^* u)$. Given a subset L of A^*, we denote by $F(L)$ $(P(L), S(L))$, the set of the words are factor (prefix, suffix) of some word in L.

A homomorphism between the free monoids A^* and Γ^* is an application $h : A^* \longrightarrow B^*$ satisfying $h(uv) = h(u)h(v)$ for all $u, v \in A^*$. Note that, the homomorphism h is completely determined by the images of letters of A in B^*, i.e, $h(a)$ for any a belong to A.
For $x, y \in A^*$, we define

$$x^{-1}y = \{z \in A^* : xz = y\} \quad \text{and} \quad xy^{-1} = \{z \in A^* : x = zy\}.$$

For subsets X, Y of A^*, this notation is extended to

$$X^{-1}Y = \bigcup_{x\in X}\bigcup_{y\in Y} x^{-1}y \text{ and } XY^{-1} = \bigcup_{x\in X}\bigcup_{y\in Y} xy^{-1}.$$

A set $X \subset A^*$ is a code if any word in X^+ can be written uniquely as a product of words in X, that is, has a unique factorisation in words in X, i.e., if for all $m, n \geq 1$ and $(x_i)_{i=1,\cdots,n}, (y_i)_{i=1,\cdots,m}$ the condition

$$x_1x_2\cdots x_n = y_1y_2\cdots y_m \text{ implies} n = m \text{ and } x_i = y_i \text{ for } i = 1, \cdots, n.$$

Any code X satisfy the Kraft inequality

$$\sum_{x\in X} (Card\,(A))^{-|x|} \leq 1$$

and for any sequence $l_1, \cdots, l_n$ of positive integers such that

$$\sum_{i=1}^{i=n} (Card\,(A))^{-l_i} \leq 1,$$

there exists a prefix code $X = \{x_1, \cdots, x_n\}$ over A such that $|x_i| = l_i$ for all $i \in \{1, \cdots, n\}$.The basic question to be asked is "When is a given subset X of A^* a variable length code?".

This was answered by Sardinas and Patterson [3]. Define recursively subsets U_n of A^* as follows:

$$\begin{cases} U_0 = X^{-1}X - \{\epsilon\} \\ U_{n+1} = U_n^{-1}X \cup X^{-1}U_n \text{ for } n \geq 0, \end{cases}$$

where ϵ denotes the identity of A^* and $X^{-1}X = \bigcup_{x\in X} x^{-1}X$. We have

- If $\epsilon \in U_n$, then X is not a variable length code;
- If $U_{n+1} = U_n$, then X is a variable length code [1, 3].

We say that, a code X is maximal if and only if for any word $z \notin X$ the set $X \cup \{z\}$ cannot be a code.

A subset X of A^* is called prefix (suffix) if $X \cap XA^+ = \emptyset$ (resp. $X \cap XA^+ = \emptyset$). A subset X of A^* is bi-prefix if it is both suffix and prefix. A code X is complete if and only if any word of A^* is a factor of some word in X^*.

For any set $X \subset A^*$, the generating function of X is

$$G_X\,(z) = \sum_{n\geq 0} g_n z^n,$$

where

$$g_n = Card\,(X \cap A^n)\,.$$

Notice that if X is a code, then [4]

$$G_{X^*} = \frac{1}{1 - G_X}.$$

The sequence $(g_n)_{n\geq 0}$ is called the length distribution of X.

Let $X, Y \subset A^*$. If X and y are disjoint, then $G_{X\cup Y} = G_X + G_Y$. Similarly, if the product of X and Y is unambiguous, that is whenever $xy = x'y'$ with $x, x' \in X, y, y' \in Y$ imply $x = x', y = y'$, then $G_{XY} = G_X G_Y$.

§3. Group Codes and Their Generating Functions

The following propositions from [2] gives a methods to construct the group codes.

Proposition 3.1 *Let G be a group and H a subgroup of G. Let $\Psi : A^* \longrightarrow G$ be a morphism. Let $X^* = \Psi^{-1}(H)$ with X the minimal generator the set X^*. We have,*

(1) *The submoinoi X^* is unitary on the right and on the left, that is whenever $xy \in X^*, x \in X^*$ then $y \in X^*$ and that is whenever $xy \in X^*, y \in X^*$ then $x \in X^*$;*

(2) *The set X is bi-prefix code;*

(3) *If Ψ is surjective, then X is a maximal bi-prefix code;*

(4) *If X is a code, then $G_X = 1 - \frac{1}{G_{X^*}}$.*

Proof (1) Suppose that $xy \in X^*, y \in X^*$, i.e, $\Psi(xy) \in H$ and $\Psi(x) \in H$, from where $\Psi(y) = (\Psi(x))^{-1}\Psi(xy)$ is in H then $y \in X^*$, so X^* is unitary on the right. On the same way, we prove that X^* is unitary on the left.

(2) As X^* is unitary on the right and on the left, then the set X is bi-prefix code.

(3) Suppose now that Ψ is surjective. If $X^* = A^*$, then $X = A$, and hence (3) is proved. Otherwise let w be any word in $A^*, w \notin X^*$. Since Ψ is surjective ($\Psi(A^*) = G$), $\Psi(w)$ is an element of the group G, so there exists $v \in A^*$ such that $\Psi(v) = (\Psi(w))^{-1}$. The words vw and wv are in X^* since $\Psi(vw) = \Psi(wv) = 1_G$. Naturally $wvw \in (X \cup \{w\})^*$, but the word wvw admits two distinct factorizations in words of $X \cup \{w\}$, that is $wvw = w(vw) = (wv)w$. The set $X \cup \{w\}$ cannot be a code, for all $w \notin Z$. Finally X is a maximal code.

(4) As $G_{X^*} = \frac{1}{1-G_X}$, then $G_X = 1 - \frac{1}{G_{X^*}}$ ($G_{X^*} \neq 0$). □

Notation 3.2 In the last case the set X is called a group code denoted by $X(G, H)_\Psi$

Example 3.3 Consider the morphism of monoids $\Psi : \{a, b\}^* \longrightarrow (\mathbb{Z}, +)$ defined by

$$\Psi(a) = 1, \Psi(b) = -1, \Psi(\epsilon) = 0.$$

And then, $\forall w \in \{a, b\}^*$ we have $\Psi(w) = |w|_a - |w|_b$.

The mapping Ψ is surjective because $\forall m \in \mathbb{Z}, \exists w \in \{a, b\}^*$ such that $\Psi(w) = m$.

In fact, we have

(1) If $m = 0$ then $w = \epsilon$;

(2) If $m > 0$ then $w = a^m$;

(3) If $m < 0$ then $w = b^{-m}$.

Let $H = \{0\}$ the trivial subgroup of $(\mathbb{Z}, +)$. Then

$$X^* = \Psi^{-1}(\{0\}) = \left\{w \in \{a,b\}^* : |w|_a = |w|_b\right\}.$$

The set X^* are the words over $\{a,b\}$ having an equal number of occurrences of a and b is a submonoid of $\{a,b\}^*$ generated by a bi-prefix code. Since any word of X^* of length $2n$ is obtained by choosing n positions among $2n$, we have

$$G_{X^*}(z) = \sum_{n\geq 0} \left(\begin{array}{c} 2n \\ n \end{array} \right) z^n.$$

Then, the sequence $\left\{ \left(\begin{array}{c} 2n \\ n \end{array} \right) \right\}_{n\geq 0}$ is the length distribution of X^*.

We show that

$$G_{X^*}(z) = \sum_{n\geq 0} \left(\begin{array}{c} 2n \\ n \end{array} \right) z^n = \left(1-4z^2\right)^{-\frac{1}{2}}.$$

In fact, we have

$$\begin{aligned}
\left(1-4z^2\right)^{-\frac{1}{2}} &= \sum_{n\geq 0} \left(\begin{array}{c} -\frac{1}{2} \\ n \end{array} \right) \left(-4z^2\right)^n \\
&= \sum_{n\geq 0} \frac{\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\left(-\frac{5}{2}\right)\cdots\left(-\frac{1}{2}-n+1\right)}{n!} \times (-1)^n \times (4)^n \times \left(z^2\right)^n \\
&= \sum_{n\geq 0} \frac{(-1)^n \times 1\times 3\times 5\cdots(2n-1)}{2^n \times n! \times n!} \times (-1)^n \times (2)^n \times (2)^n \times n! \times \left(z^2\right)^n \\
&= \sum_{n\geq 0} \frac{(2n)!}{n!\times n!} \left(z^2\right)^n \quad (\text{note that } (2)^n \times n! = 2\times 4\times 6\cdots\times 2n) \\
&= \sum_{n\geq 0} \left(\begin{array}{c} 2n \\ n \end{array} \right) z^{2n}.
\end{aligned}$$

As

$$G_{X^*} = \left(1-4z^2\right)^{-\frac{1}{2}},$$

then

$$G_X(z) = \left(1 - \frac{1}{G_{X^*}}\right)(z) = 1 - \left(1-4z^2\right)^{\frac{1}{2}}.$$

We get that

$$\begin{aligned}
G_X(z) &= 1-\left(1-4z^2\right)^{\frac{1}{2}} = 1-\sum_{n\geq 0}\left(\begin{array}{c}\frac{1}{2}\\ n\end{array}\right)\left(-4z^2\right)^n \\
&= 1-\left(1+\sum_{n\geq 1}\left(\begin{array}{c}\frac{1}{2}\\ n\end{array}\right)\left(-4z^2\right)^n\right) = -\sum_{n\geq 1}\left(\begin{array}{c}\frac{1}{2}\\ n\end{array}\right)\left(-4z^2\right)^n \\
&= -\sum_{n\geq 1}\frac{\left(\frac{1}{2}\right)\left(\frac{-1}{2}\right)\left(\frac{-3}{2}\right)\left(\frac{-5}{2}\right)\cdots\left(\frac{-1}{2}-n+1\right)}{n!}\times(-1)^n\times(4)^n\times\left(z^2\right)^n \\
&= -\sum_{n\geq 1}\frac{\left(\frac{1}{2}\right)\left(\frac{-1}{2}\right)\left(\frac{-3}{2}\right)\left(\frac{-5}{2}\right)\cdots\left(\frac{-(2n-3)}{2}\right)}{n!}\times(-1)^n\times(2)^n\times(2)^n\times z^{2n} \\
&= -\sum_{n\geq 1}\frac{(-1)^n\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{3}{2}\right)\left(\frac{5}{2}\right)\cdots\left(\frac{(2n-3)}{2}\right)}{n!}\times(-1)^n\times(2)^n\times(2)^n\times z^{2n} \\
&= \sum_{n\geq 1}\frac{(1)(3)(5)\cdots(2n-3)}{n!\times n!}\times n!\times(2)^n\times z^{2n} \\
&= \sum_{n\geq 1}\frac{(2n)!}{(2n-1)\,n!\times n!}z^{2n} = \sum_{n\geq 1}\frac{2}{n}\left(\begin{array}{c}2n-2\\ n-1\end{array}\right)z^{2n}.
\end{aligned}$$

Finally the generating function of X is

$$G_X(z) = \sum_{n\geq 1}\frac{2}{n}\left(\begin{array}{c}2n-2\\ n-1\end{array}\right)z^{2n}.$$

Example 3.4 Let $\Psi : A^* \longrightarrow (\mathbb{Z}/n\mathbb{Z}, \oplus)$ the morphism of monoids defined by

$\Psi(a) = \overline{1}$ *for all* $a \in A$, *and* $\Psi(\epsilon) = \overline{0}$. *And consequently,* $\forall w \in A^* : \Psi(w) = |w|\, mod(n)$.

We have the mapping Ψ is surjective because $\forall \overline{m} \in \mathbb{Z}/n\mathbb{Z}$, the word $w = \sigma^m \in A^*$, for all $\sigma \in A$, satisfies the condition $\Psi(\sigma^m) = \overline{m}$. And if $X^* = \Psi^{-1}(\{\overline{0}\}) = \{w \in A^* : |w| \equiv 0 \mathrm{mod}\ \ (n)\}$ then, $X = A^n$. We have

$$G_X(z) = \sum_{n\geq 0} g_n z^n,$$

where $g_n = Card(A^n) = (Card(A))^n$, then

$$G_X(z) = \sum_{n\geq 0}(Card(A))^n z^n = \frac{1}{1-(Card(A))^n z}.$$

The generating series of X^* is

$$G_{X^*} = \frac{1}{1-G_X} = \frac{1}{1-\frac{1}{1-(Card(A))^n z}} = \frac{(Card(A))^n z - 1}{(Card(A))^n z}.$$

Proposition 3.5 *Let $X(G,H)_\Psi$ be an arbitrary group code. If the morphism Ψ is surjective, then $X(G,H)_\Psi$ is complete.*

Proof We show that any word of A^* is a factor of some word in X^*. Let $w \in A^*$, the word w is a factor of $uwv \in X^*$, where $\Psi(u) = (\Psi(w))^{-1}$ and $\Psi(v) = \Psi(\epsilon) = 1_G$. Consequently we obtain $A^* = F(X^*)$. □

Example 3.6 Let $\Psi : \{a,b\}^* \longrightarrow (\mathbb{Z},+)$ defined by: $\Psi(a) = 1, \Psi(b) = -1, \Psi(\epsilon) = 0$. And then, $\forall w \in \{a,b\}^*$ we have $\Psi(w) = |w|_a - |w|_b$.

The morphism Ψ is surjective. In fact, let $X^* = \Psi^{-1}(\{0\}) = \left\{w \in \{a,b\}^* : |w|_a = |w|_b\right\}$. Any word w of $\{a,b\}^*$ is a factor of $uwv \in X^*$, where $\Psi(u) = (\Psi(w))^{-1} = -(|w|_a - |w|_b) = |w|_b - |w|_a$ and $\Psi(v) = \Psi(\epsilon) = 1_G$. otherwise the word is the factor of $uwv \in X^*$, where $|v|_a = |w|_b, |v|_b = |w|_a$ and for example $v = ab$. In this case, we have $\Psi(uwv) = 0$.

Example 3.7 Let $\Psi : A^* \longrightarrow (\mathbb{Z}/n\mathbb{Z}, \oplus)$ the morphism of monoids defined by $\Psi(a) = \overline{1}$ for all $a \in A$ and $\Psi(\epsilon) = \overline{0}$. Consequently, $\forall w \in A^* : \Psi(w) = |w| \, mod(n)$. In fact, the morphism Ψ is surjective. Let $X^* = \Psi^{-1}(\left\{\overline{0}\right\}) = \{w \in A^* : |w| \equiv 0 \text{mod } (n)\}$. Any word w of A^* is a factor of $uwv \in X^*$, where $\Psi(u) = (\Psi(w))^{-1} = -(|w| \, mod(n))$ and $v = \epsilon$.

§4. Conclusion

In this work, we have calculated the generating functions of some group codes and its stars.

References

[1] D. Clelia, Some conjectures on codes, arXiv preprint *arXiv*: 1611.04580 (2016).

[2] P. Dominique and G. Rindone, On syntactic groups, *Bulletin of the Belgian Mathematical Society Simon Stevin*, 3 (2004), 749-760.

[3] J. Falucskai, On equivalence of two tests for codes, *Acta Mathematica Academiae Paedagogicae Nyíregyháziensis*, 24.2 (2008) 249-256.

[4] B. Frédérique, Generating functions of circular codes, *Advances in Applied Mathematics*, 22.1 (1999) 1-24.

[5] N. Ghadbane and D. Mihoubi, A construction of some group codes, *International Journal of Electronics and Information Engineering*, 4 (2) (2016) 55-59.

[6] A. Ginzburg, *Algebraic Theory of Automata*, Academic Press, New York, (1968).

[7] R. Giuseppina, Construction d'une famille de codes associes a certains groupes finis, *Theoretical Computer Science*, 54 (1987) 165-179.

[8] B. Jean and P. Dominique, *Theory of Codes*, Academic Press (1985).

[9] N. Jean, Completing prefix codes in submonoids, *Theoretical Computer Science*, 356 (2006) 245-254.

[10] B. Jean and al., Recent results on syntactic groups of prefix codes, *European Journal of Combinatorics*, 33 (7) (2012) 1386-1401.

[11] S. Marcel-Paul, On a question concerning certain free submonoids, *Journal of Combinatorial Theory*, (1966) 437-442.