



HAL
open science

Balance properties on Christoffel words and applications

Lama Tarsissi

► **To cite this version:**

Lama Tarsissi. Balance properties on Christoffel words and applications. General Mathematics [math.GM]. Université Grenoble Alpes, 2017. English. NNT : 2017GREAM097 . tel-01885914

HAL Id: tel-01885914

<https://theses.hal.science/tel-01885914>

Submitted on 2 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Mathématiques et Informatique**

Arrêté ministériel : 25 mai 2016

Présentée par

Lama Tarsissi

Thèse dirigée par **Laurent VUILLON, USMB**

préparée au sein du **Laboratoire de mathématiques LAMA, Université Savoie Mont Blanc**
et de l'**École Doctorale Mathématiques, Sciences et Technologie de l'Information, Informatique**

Balance properties on Christoffel words and applications.

Thèse soutenue publiquement le **24 Novembre 2017**,
devant le jury composé de :

Monsieur Laurent VUILLON

Professeur, Université Savoie Mont Blanc, Directeur de thèse

Monsieur Jean-Pierre BOREL

Professeur, Université de Limoge, Examineur et Président du Jury

Monsieur Michel RIGO

Professeur, Université de Liège, Rapporteur

Monsieur Jean-Marc FÉDOU

Professeur, Université de Nice Sophia-Antipolis, Rapporteur

Monsieur Simone RINALDI

Professeur, Université de Sienne, Examineur

Monsieur Julien CASSAIGNE

Chargé de Recherche, Institut de Mathématiques de Marseille, Examineur



Contents

1	Some combinatorics on words	9
1.1	Notation and preliminaries	9
1.2	Lyndon words	11
1.3	Standard Sturmian Words	13
1.4	Christoffel words	15
1.4.1	Standard factorization	18
1.4.2	Christoffel tree	21
1.5	Balance property	23
1.5.1	Balance matrix	24
1.5.2	Properties of the balance matrix	29
1.5.3	Christoffel words with a and b not coprime	34
2	Algebraic view	37
2.1	Algebraic definition	37
2.2	More about Christoffel words	44
2.2.1	Continued fractions	45
2.2.2	Stern-Brocot tree	48
2.2.3	Second construction of the balance test matrix	64
3	Second order balanced matrix	69
3.1	Second order balance matrix	70
3.2	Recursive construction of the second order balance matrix	72
3.2.1	Properties of the matrix $U_{\frac{a}{b}}$	73
3.2.2	General form of the second order balance matrix	74
3.3	The construction of $U_{\frac{a}{b}}$	76
3.3.1	The trivial cases : $z \in \{0, 1\}$	76
3.3.2	The general case : $z \geq 2$	77
4	Some results	89
4.1	Second order balance of $C(\frac{a}{b})$	89

4.1.1	Refinement of the value of $\delta^2(\frac{a}{b})$ using the continued fraction	92
4.1.2	Fibonacci sequence	97
4.2	Abelian vectors	101
5	Symmetric standard factorization	111
5.1	Symmetric standard factorization	111
5.1.1	Geometrical examples	120
6	Synchronization of Christoffel words	127
6.1	Synchronization of Christoffel words	128
6.2	Vertical invariant	130
6.3	Seeds for two generators	132
6.4	Horizontal invariant	135
6.5	Case of three generators	136
6.5.1	Generators not relatively co prime	139
6.6	Relation between $I_{n,k}$ and $I_{h_{g_i}}$	141
6.7	Equal generators	147
6.8	Distinct generators	148
6.8.1	Fraenkel's seed	149
6.9	Relation between 2 and 3 generators	151
7	Convexity and digital convex polyominoes	157
7.1	Introduction	158
7.1.1	Discrete geometry and digital space	159
7.2	Theoretical results	163
7.2.1	Perturbations on the WN paths	164
7.2.2	Definition of the <i>split</i> operator	168
7.2.3	Commutativity of the <i>split</i> operator	169

Remerciement

En premier lieu, je remercie mon directeur de thèse Laurent Vuillon qui a été toujours là pendant ces trois années de travail et de recherche. On a bien collaboré ensemble et discuté pour trouver des résultats. Tu as pu trouver les idées et les sujets qui se synchronisent avec mes compétences. Tu étais pour moi un grand support. Je te remercie bien à la bonne relation professionnelle qui m'a donné une grande confiance pendant ces trois années. Je remercie aussi le directeur du laboratoire Philippe Briand.

Également, je remercie Xavier Provençal, qui a été toujours là pour m'aider à améliorer mes codes et à me donner des conseils. Je te remercie pour tout ce que tu m'as appris. En plus, j'ai eu la chance de collaborer, échanger des idées et écrire un article avec les chercheurs italiens qui visitent annuellement notre Labo: Andrea Frosini, Simone Rinaldi.

Pareillement, je remercie mes collègues au labo avec qui j'ai passé des jolis moments, on a bien discuté sur des différents sujets et ils m'ont bien aidé à m'intégrer au labo. Je remercie Rodolphe Lepigre qui a répondu à toutes mes questions. Je n'oublie pas Pierre Villemot qui a enseigné en classes parallèles avec moi et on a passé du temps à discuter et à faire des formations ensemble. Marion Foare et Florian Patout merci bien pour cette dernière année et pour votre aide pendant le déménagement. Je n'oublie pas mes collègues Rodiguo Dorantes, Clovis Eberhart, Rémy Nguyen, Charlotte Perrin et Paul el Hilany. Finalement, je remercie bien Bilal el Takki, Joseph El Maalouf et Ralph Lteif mes collègues Libanais qui leur présence à Chambéry m'a aidée à sentir que je suis chez moi et surtout pendant les pauses cafés.

Hors du labo, j'ai bien rencontré des amis qui ont facilité et simplifié ma présence à l'étranger. Je tente à remercier Ranime el Nabouch ma collocataire, Mohammad Ibrahim, Mohannad Dbeiss, Lauréna et Pia Dally qui sont devenus une partie de ma famille. Je n'oublie pas Abbas Abbas qui a été le premier à m'aider dès mon arrivée en France et à me loger. Nissrine Raad, Hadi Ballout, Ali Bourgi merci pour les jolis moments qu'on a passé ensemble.

Finalement, je passe au Liban, pour remercier Zahraa Haydar et son frère Jawad qui avec leur famille étaient toujours là pour me motiver et m'encourager.

Riad Bou Hadir le vrai ami depuis ma vie à l'étranger à Dubai, merci bien pour tes conseils et ton support. Rayane Fakih et Mario Ibrahim mes cousins à Grenoble vous étiez le symbole de ma famille ici en France. La famille Baliki (Roger, Mirna, Daniel et Christian), c'est avec gratitude que je vous ai rencontré et vous êtes devenus ma famille Canadienne.

Comme le dessert est toujours servi à la fin. Pour cela j'ai laissé le grand remerciement pour la fin où les mots ne sont pas suffisants pour exprimer ma reconnaissance. C'est pour cette raison que je dédie mon travail et ma réussite à mes parents. Mon père Youssef Tarsissi, ma mère Sana Fakih, mon frère Jad Tarsissi et ma soeur Jala Tarsissi sans votre support, votre motivation et votre croyance en moi je n'aurai pas pu le faire. Pour cela je vous dis merci pour votre patience contre la distance et pour tout ce que vous avez fait et supporté pour m'aider à obtenir mon diplôme.

Résumé

De nombreux chercheurs se sont intéressés à la combinatoire des mots aussi bien d'un point de vue théorique que pratique. Pendant plus de 100 ans de recherche, de nombreuses familles de mots ont été découvertes, certaines sont infinies et d'autres sont finies. Dans cette thèse, on s'intéresse aux mots de Christoffel. On aborde aussi les mots de Lyndon et les mots Sturmien standards. De plus, nous donnons de nombreuses propriétés sur les mots de Christoffel et on approfondit l'étude de la notion d'équilibre. Il est connu que les mots de Christoffel sont des mots équilibrés sur un alphabet binaire et sont formés par la discrétisation de segments de droite de pente rationnelle. Les mots de Christoffel sont aussi retrouvés dans l'étude de la synchronisation de k processus à l'aide de k mots équilibrés. Pour $k = 2$, on retombe sur les mots de Christoffel, tandis que pour $k > 2$, la situation est plus compliquée et nous amène à la conjecture de Fraenkel qui est ouverte depuis plus de 40 ans. Comme c'est difficile d'atteindre cette conjecture, alors nous avons cherché à construire des outils qui nous aide à s'approcher de cette conjecture. On introduit ainsi la matrice d'équilibre B_w où w est un mot de Christoffel et la valeur maximale de cette matrice est l'ordre d'équilibre du mot binaire utilisé. Comme les mots de Christoffel sont équilibrés alors la valeur maximale dans ce cas-là sera égale à 1 et chaque ligne de cette matrice sera formée des mots binaires. Cela nous pousse à tester de nouveau l'ordre d'équilibre de chaque mot obtenu et une nouvelle matrice est obtenue qui s'appelle matrice d'équilibre du second ordre. Cette matrice admet plusieurs propriétés de symétrie et a une forme particulière comme on est capable de la partager en 9 blocs où c'est suffisant de savoir 3 parmi eux pour construire le reste. Ces trois blocs correspondent à des matrices de mots de Christoffel qui se trouvent dans des niveaux plus proches de la racine de l'arbre des mots de Christoffel. La valeur maximale de cette nouvelle matrice U_w est appelée équilibre du second ordre. En regardant les chemins qui minimisent cette valeur tout au long de l'arbre, on remarque que le chemin suivi par les fractions obtenues du rapport des nombres consécutifs de la suite de Fibonacci, appelé chemin de Zig-zag est l'un des chemins minimaux. On retrouve ces chemins géométriquement sur le chemin de Christoffel en introduisant une nouvelle factorisation pour les mots de Christoffel appelée la factorisation standard symétrique. Nous avons, également, pu trouver une relation directe entre la matrice U_w et le mot de Christoffel initial sans passer par la matrice B_w et cela en étudiant l'ensemble des vecteurs abéliens associés. Tout ce travail nous a permis de réfléchir au sujet initial qui est la synchronisation de k mots équilibrés. Ainsi, pour le cas de 3 générateurs, nous avons pu étudier

tous les cas possibles de la synchronisation et une discussion bien détaillée est faite en utilisant un nouvel élément appelé la graine qui est la première colonne de la matrice de synchronisation. La matrice du second ordre d'équilibre, avec toutes ses propriétés va être un bon outil pour étudier la synchronisation de k générateurs et cela constitue mon projet de recherche dans le futur. Nous avons aussi utilisé toutes nos connaissances autour des mots de Christoffel pour avancer dans la reconstruction de polyominoes convexes. Comme le contour d'un tel polyomino est formé des mots de Christoffel de pentes décroissantes, on a introduit un nouvel opérateur qui modifie ce chemin tout en gardant la décroissance des pentes c'est-à-dire en conservant la convexité qui est un premier pas vers la reconstruction.

Abstract

Many researchers have been interested in studying Combinatorics on Words in theoretical and practical points of view. Many families of words appeared during these years of research some of them are infinite and others are finite. In this thesis, we are interested in Christoffel words and we introduce the Lyndon words and Standard Sturmian words. We give numerous properties for this type of words and we stress on the main one which is the order of balancedness. Well, it is known that Christoffel words are balanced words on two letters alphabet, where these words are exactly the discretization of line segments of rational slope. Christoffel words are considered also in the topic of synchronization of k process by a word on a k letter alphabet with a balance property in each letter. For $k = 2$, we retrieve the usual Christoffel words. While for $k > 2$, the situation is more complicated and lead to the Fraenkel's conjecture that is an open conjecture for more than 40 years. Since it is not easy to solve this conjecture, we were interested in finding some tools that get us close to this conjecture. A balance matrix B_w is introduced, where w is a Christoffel word, and the maximal value of this matrix is the order of balancedness of the binary word. Since Christoffel words are one balanced then the maximal value obtained in this matrix is equal to 1 and all the rows of this matrix is made of binary words. Testing again the balancedness of these rows, a new matrix arises, called second order balance matrix. This matrix has lot of characteristics and many symmetries and specially the way it is constructed since it is made of 9 blocks where three of them belong to some particular Christoffel words appearing in some levels closer to the root of the Christoffel tree. The maximal value of this matrix is called the second order of balancedness for Christoffel words. From this matrix and this new order of balancedness, we were able to show that the path followed by the fractions obtained from the ratio of the consecutive elements of Fibonacci sequence is a minimal path in the growth of this second order. In addition to that, these blocks are geometrically found on the Christoffel path, by introducing a new factorization for the Christoffel words, called Symmetric standard factorization. Similarly, we worked on finding a direct relation between the second order balance matrix U_w and the initial Christoffel word without passing by the balance matrix B_w but by studying the set of factors of abelian vectors. All this work allow us to think about the initial topic of research which is the synchronization of k balanced words. A complete study for the case $k = 3$ is given and we have discussed all the possible sub-cases for the synchronization by giving its seed, which is the starting column of the synchronized matrix. The second order balance matrix, with all its

properties and decompositions form a good tool to study the synchronization for k generators that will be my future project of research. We have tried to use all the knowledge we apply them on the reconstruction of digital convex polyominoes. Since the boundary word of the digital convex polyominoe is made of Christoffel words with decreasing slopes. Hence we introduce a split operator that respects the decreasing order of the slopes and therefore the convexity is always conserved that is the first step toward the reconstruction.

Introduction

A solar year is composed of 365.25 days where the moon appears fully illuminated each 29.53 days. Doing the calculations, we can note that a solar year can have 12 full moons or 13 full moons. Scientists have realized that by assigning the number 0 to the year with 12 moons and the number 1 to the year with 13 moons, the sequence of numbers obtained is a balanced binary word. Note that a binary word is balanced if for any two factors of the same length, the number of occurrences of the letter 1 in both words is either the same or different by one occurrence [5, 53]. This is a direct application for the combinatorics on words, the field that has been widely studied in recent years. We can observe that balancedness is something that we can find in our day to day life without notice. Indeed, most of this theory is developed implicitly as a tool to attack several problems which arise from Computer Science. This thesis deals with finite binary words called Christoffel words. Those words are fundamental objects in combinatorics which share many geometric, algebraic and arithmetic properties [5, 14]. Among them and in discrete geometry, Christoffel words are coding of a discrete segment on a grid from $(0, 0)$ to (a, b) [40]. Christoffel words have an algebraic interpretation using Cayley graph and two generators $g_1 = a, g_2 = b$ over $\mathbb{Z}/n\mathbb{Z}$, with $n = a + b$ [5, 14]. The study of these words began in 1873 by Elwin Christoffel [14] and the community of combinatorics on words. They were able to give new applications to Christoffel words because these words are balanced. In fact, we are able to find many studies done by Elwin Christoffel [25] and Reutenauer [5] to find characterizations of these words.

Finding balanced words was always the purpose of many mathematicians like Etian Altman [2], Pascal Hubert [32] and Robert Tijdeman [51]. In this thesis, we try to refine this balancedness concept by studying a second order of balancedness in the interest of classifying the Christoffel words by their order of balancedness and understanding deeply the structure of balanced words. Hence, we start by introducing a matrix called the *balanced matrix* of a binary word that tests its order of balancedness. As we said before, since we will be dealing only with Christoffel words, and since those words are 1-balanced, the matrix obtained for this family of words, is composed of binary words which pushed us to investigate another order of balancedness. Therefore, we propose a second order balance for 1-balanced binary words and show some of its properties. The second order balance can be summarized as the maximal balancedness of the words that describes the number of occurrences of a given letter in all factors of a given length. This allows a classification among 1-balanced words. In another words, we can say that for each given length j , we can find, in a balanced word v , blocks containing k_j or

k_{j+1} the letter 1. The second order balance is the balancedness over the words coding the successive positions of the blocks containing k_j or k_{j+1} the letter 1 in the balanced word v . Hence, we are working for a given value j on the repartition of these blocks in a binary balanced word. This second order balance is computed via a matrix called the *Second order balance matrix*. The properties of this matrix are obtained by checking the composition of the Christoffel words and considering some of its particular rows.

In another perspective, since doing things on time is universal, nearly every activity requires synchronized timing to operate at peak levels. During an airplane's departure, on signal lights at an intersection, in robotics or in any factory's assembly line, synchronization is needed, otherwise everything would happen simultaneously and in disarray. Therefore synchronization is an important concept to study. In addition to that, many algorithms in discrete geometry and computer imagery deal with plotting segments on a computer screen and thus Christoffel words are also fundamental objects for this topic [40]. In parallel, many computer scientists work on synchronization of processes. They try to design optimization algorithms in order to balance the charge between more than two servers and the optimal solution. In the case of two servers, the word obtained due to the synchronization is given either by aperiodic balanced words named the Sturmian words or by periodic balanced words named Christoffel words repeated periodically [2]. The question that comes to the mind immediately is the following: What will happen if we have three servers or more? It is clear that the problem would become more complicated and would lead to the study of balanced words on a three letter alphabet or more. We can see that a good example of synchronization is the one given by another community that used the Beatty sequences in order to give the positions of each of the letters in balanced words. For more details of this approach, see the works of Fraenkel [24, 25]. In fact, the usual Christoffel word can be seen as a synchronization using two generators like the work done by Reutenauer [38]. Aviezri Fraenkel, in his papers [24, 25, 50, 52] has mentioned many ways to prove that the unique solution for balanced words on k -letters alphabet with pairwise distinct frequencies of letter, where $3 \leq k \leq 6$, is the sequence of generators of the form: 2^{k-1} called **Fraenkel's conjecture** for all $k \geq 3$. Many studies use either balanced words or Beatty sequences like Ronald Graham [30], nevertheless Christoffel words can give a better insight on Fraenkel's conjecture. These studies were able to prove the conjecture for $k = 7$, using approaches which do not have similar methods to the one used by Fraenkel. In particular, Fraenkel stated its famous conjecture from 40 years (translated in combinatorics on words vocabulary) that the unique solution (up to permutation of letters) for balanced words on k -letters alphabet with pairwise distinct frequencies of letters is given

by repeating periodically the Fraenkel word $Fr_k = Fr_{k-1}kFr_{k-1}$ for $k > 2$ and $Fr_2 = 121$. For instance, we have $Fr_3 = 1213121$ and $Fr_4 = 121312141213121$.

The formulation of the original statement of the Fraenkel conjecture is the following:

Conjecture 1 *There exists a unique periodic balanced words (up to permutation of letters) on k letters with pairwise distinct frequencies of letters and the k frequencies are exactly $\frac{2^k}{2^{k+1}-1}, \frac{2}{2^{k+1}-1}, \frac{1}{2^{k+1}-1}$.*

Example 1 *The frequencies of letters in $Fr_3 = 1213121$ are given by $\frac{4}{7}, \frac{2}{7}, \frac{1}{7}$ and in $Fr_4 = 121312141213121$ by $\frac{8}{15}, \frac{4}{15}, \frac{2}{15}, \frac{1}{15}$.*

This conjecture is proved for $k = 3, \dots, 7$ in the following works [2, 46, 52] and the approach uses extensively the properties of Beatty sequences [30, 47], while it is still open for $k > 7$. Nevertheless, the study of balanced words on k letters for $k > 2$ is of course a topic in combinatorics on words. For example Tijdeman's and Hubert's works use Sturmian sequences in order to construct balanced words [32, 50, 53]. We remark that the work of Hubert is a type of restatement of Graham's work in combinatorics on words vocabulary.

Now, in another perspective, we need to investigate a new discrete tool in the study of the synchronization of Christoffel words. This tool was introduced in the work of G. Paquin and L. Vuillon in 2007. In fact, they were able to give a link between Fraenkel's conjecture and Christoffel word construction by introducing the column invariant on the synchronization of k Christoffel words [38]. In parallel, G. Paquin and C. Reutenauer wrote an article about the synchronization of k Christoffel words on k letters and an extra symbol $*$ and found many properties of partial synchronization of Christoffel words (the $*$ materialize an extra gap letter in the synchronization) and formulate many properties from the Beatty sequences point of view to the combinatorics on words point of view [39]. Based on this link and using the properties of the second order balance matrix, we aim in this thesis to give a new tool and a new perspective to discuss the synchronization of Christoffel words. Since it is difficult to reach the Fraenkel's conjecture during those three years of research, we let the second order balance matrix and its properties be a good tool to get closer to this conjecture.

Organization of the thesis

Chapter I: Combinatorics on words is a wide domain of research. Many family of words were discovered and studied in the last decades. Therefore, in this chap-

ter, we start by introducing the common literature used all over these years. We define what is an alphabet, the set of letters used to form a word. We give the definition of palindromes, suffixes and prefixes. To be able to reach the goal of this thesis, we have to illuminate on some properties and mention some families of words over the binary alphabet $A = \{0, 1\}$. Hence, we start by introducing the lexicographic order to be able to define the Lyndon words. After that, we pass to the family of Standard strumian words that is the family of infinite words used for the discretization of a line with irrational slope starting from the origin. In the finite case, when the slope is a rational number of the form a/b , we get the family of Christoffel words. These Christoffel words can be written in a particular way using a specific factorization called *Standard factorization* where this factorization leads to construct the Christoffel tree that is the tree containing all the Christoffel words and starting with the word $w = 01 = (0, 1)$ at the top of the tree. The balancedness of the words formed by the last two families is a special characteristic that will be the main topic of this thesis. The balance property allows us to define a balance matrix that gives the order of balancedness, noted δ , for any binary word. This value is equal to 1 for the particular case of Christoffel and Strumian words since they are balanced words. Many properties of this matrix are exposed in this chapter. Several examples are given in this chapter to explain those properties and relations like the complementary rows concept and the case of non primitive Christoffel words.

Chapter II Christoffel words can be seen with another perspective and can have an equivalent algebraic definition. In this chapter, we present in algebraic way the Christoffel words, where we determine the position of the decreasing positions for each word using the link to the Cayley graph and the theorem of Paquin-Reutenauer [38]. The Christoffel word is the discretization of the line segment of rational slope a/b , hence we can't talk about this family without mentioning the development of continued fractions of its slope. Therefore, we can also see in this chapter the definition of the continued fractions and its link with the Christoffel words. Rational numbers push us to think immediately to the tree that contains all the rational numbers with numerator and denominator relatively prime. This tree, with all its properties, is revealed in this chapter and the most valuable part is the link given between this tree and the Christoffel tree. Finally, using the algebraic definition of the Christoffel words, we give for the balance matrix B_w a second construction based on the set of decreasing positions of the word defined at the beginning of this chapter.

Chapter III The balance matrix B_w defined in two different ways in Chapters 1 and 2 is used in this thesis over a binary alphabet. In particular, when we restrict

the work to Christoffel words $w = C(\frac{a}{b})$ that are balanced words, we have that δ_w which is the order of balancedness and the maximal value of this matrix is equal to 1. In this case, we remark that each row of B_w is composed of a binary word. This pushes us to think and try to study again the balancedness of each row of B_w . A second order balance matrix U_w is constructed from the balance matrices of each row of the initial balance matrix B_w . We let the second order of balancedness be the maximal value of U_w . After computing this new matrix, we realize that this matrix has many symmetries and properties that are shown in this chapter. Much more, a recursive construction using the Stern-Brocot tree is obtained while comparing the result given by this second order balance matrix. We are able to show that each matrix U_w can be decomposed into 9 blocks, and due to the symmetries already shown it is enough for us to study the construction of three of them. Those three blocks, and depending on the position of the slope of w on the Stern-Brocot tree, use three or four previous and smaller matrices belonging to some fractions situated in the Stern-Brocot tree few levels before a/b . We show in detail how each block is constructed using the algebraic definition of the Christoffel words and the continued fractions expansion of these smaller rational numbers.

Chapter IV: Two important properties are deduced from the second order balance matrix. By contributing to each fraction $\frac{a}{b}$ on the Stern-Brocot tree the value corresponding to $\delta^2(C(\frac{a}{b}))$, we remark that using the recursion used to construct the second order balance matrix, this value is obtained recursively. Hence, we show how can we obtain this value by passing from level to level in the tree and what are the boundaries of this δ^2 . We also try to refine this result by linking it to the development in continued fraction of each rational number. A very interesting result appears by following the path formed by the fractions obtained from the ratio of two consecutive numbers of the Fibonacci sequence. First of all, this path is particular because it has a zig-zag form since the expansion in continued fractions is of the form $[0, 1, 1, 1, \dots]$. Secondly, this path converges to the golden number φ and is one of the paths that minimizes the growth of δ^2 in this tree and achieves the minimal value for each level of the tree. A second point is discussed in this chapter, in fact, while constructing the second order balance matrix U_w , we had to pass by the balance matrix B_w and study the balancedness of each binary word obtained in this matrix. A trivial question arises after computing this matrix, are we able to get these entries immediately from the Christoffel word itself without passing by its balance matrix? In Chapter 5, we show that the answer to this question is positive and using the abelian vectors of the Christoffel word w , we are able to determine the exact value of each element of this matrix.

Chapter V Not too far from what we said before, in this chapter and as we

did with the Christoffel words, we give a second interpretation for the second order balance matrix. In fact, in Chapter 3, we have shown the algebraic construction of the matrix U_w , while in this chapter, we explain the geometric point of view and construction of this matrix on the Christoffel path. To be able to manifest this part, we define a new type of factorization for Christoffel words called the *Symmetric standard factorization* that is deduced from the standard factorization defined in Chapter 1 and the palindrome concept. The Christoffel path has two particular points that are mentioned in Chapter 1. Their utility starts to appear from this chapter since this new factorization uses these two points and helps us to see all the blocks of the second order balance matrix on the Christoffel path.

Chapter VI The initial problem of studies in this thesis is the synchronization of Christoffel words. We let two Christoffel words of same length n over the binary alphabets $\{1, *\}$ and $\{2, *\}$ where $*$ represents the increasing step or position in each Christoffel word. We mean with synchronization, when certain circular permutations are applied to each word in such a way in each column we have no $*$ or no intersecting values. The first column in the synchronized form, where the synchronized form is constructed from the multiples of each generator modulo n , is called seed. In fact, we know that two generators are always synchronized and have the following seed: $S = \begin{pmatrix} n-1 \\ 0 \end{pmatrix}$. In this chapter, we generalize the work and we introduce a new constant called $I_{n,k}$ that helps us to study all the possible cases with three generators. We give the general form of the seed when we have equal generators and in particular, for the case of three generators, we prove

that depending on the parity of n we can get either the seed: $S = \begin{pmatrix} n-1 \\ \frac{n-1}{2} \\ 0 \end{pmatrix}$

if n is odd or $S = \begin{pmatrix} n-1 \\ \frac{n-2}{2} \\ 0 \end{pmatrix}$ if n is even if we have two equal generators and

one distinct. Indeed, the row starting with 0 corresponds always to the distinct generator. While when the generators are all distinct, we came back to Fraenkel's conjecture that was proved for $k \leq 7$, hence since we know that for $k = 3$ the generators are synchronized, we had only to give the form of the seed. Well, we prove that for any number of generators the seed of the generators of the form

$1, \dots, 2^{k-1}$ is always of the form: $S = \begin{pmatrix} \frac{I_{n,k}}{k} \\ \vdots \\ \frac{I_{n,k}}{k} \end{pmatrix}$. Due to the lack of time, and since

it is a difficult conjecture, we were not able to go further in this work to study the

synchronization for 4 generators or to see how to prove the Fraenkel's conjecture for $k > 7$. But on the other hand, and using the second order balance matrix, we can read several information that can be a tool leading us in further works to solve this conjecture.

Chapter VII In this chapter, we apply the Christoffel words and Lyndon words in discrete tomography. Working in a discrete and digital plane \mathbb{Z}^2 , a convex path is said 4 connected if we can move from a digital square to another moving in 4 directions, up, down, left or right. While the path is said to be 8 connected if in addition to these four directions, we are able to move in the four diagonal directions. A polyominoe is a 4 connected path with no holes, which means that the complementary of this polyominoe in \mathbb{Z}^2 is also convex. The boundary of a digital convex polyominoe is a word over an alphabet of cardinal 4 basically noted $A = \{0, 1, \bar{0}, \bar{1}\}$. Several studies appear to get a reconstruction algorithm for digital convex discrete sets. In fact the recent result by J.O.Lachaud and al. have shown that if the boundary word can be factorized using the Lyndon factorization where each factor is a Christoffel word, hence it is a digital convex polyominoe. In other words, the boundary word is formed of Christoffel words with decreasing slopes. In addition to that, in this chapter, we provide some local properties that a boundary word has to fulfill in order to allow a single point modification and to preserve the convexity. As we have mentioned before, the Christoffel path has two particular points. The first one is the closest point of the path with respect to the line segment and the other is the furthest point of the path with respect to the line segment. In fact, the point needed for this modification is the furthest one, in fact, at this position we do a split operation by exchanging the 0 by 1 and the 1 by 0. With this operation, we are able to show that digital convexity is conserved so that the slopes are always in a decreasing order.

During all this work, all the computations and tests were verified and checked using the software Sage based on Python. Therefore, in each chapter, you can find the code used to test each result followed by an example of application.

Chapter 1

Some combinatorics on words

1.1 Notation and preliminaries

Combinatorics on words are related to several branches of mathematics and considered as an interesting topic of studies. We recall some classical notations and concepts on words basically introduced and defined in [33].

An **Alphabet** A is a finite set of symbols such that its elements are called **letters**. We call a **word** w over an alphabet A , a sequence of letters over this alphabet. In another way, a word w is obtained by *concatenating* letters of the alphabet A and we write $w \in A^*$, where A^* represents the set of all the words formed by the alphabet A .

A^* is considered as a free monoïd where its identity element is noted ϵ and is called the **empty word**. We let A^+ represents the set of all non-empty words over the alphabet A , i.e $A^+ = A^* \setminus \epsilon$. We denote $w^n \in A^*$ the concatenation of the word w , n times, which means: $w^n = www \cdots w$ where w is repeated n times with the convention that $w^0 = \epsilon$. A word w is said **primitive** if it is not the power of a nonempty word.

Example 2 Consider the alphabet $A = \{0, 1\}$, the word $w = 00100101$ belongs to A^* .

Let $w \in A^*$, we represent by n the **length** of the word, which is the number of letters of the word w denoted by $n = |w|$, where $|\epsilon| = 0$. For all $\ell \in A$, $|w|_\ell$ denotes the number of occurrences of the letter ℓ in the word w and we write:

$$|w| = \sum_{\ell \in A} |w|_\ell.$$

Example 3 Consider the word $w = 1001010010$, we have $|w| = 10$, $|w|_0 = 6$, and $|w|_1 = 4$.

Let $p, f, s \in A^*$ such that $w = pfs$, we say that p is a **prefix** of w , f is a **factor** of w and s is a **suffix** of w . Since any of the words p, f, s might be ϵ , then prefixes and suffixes are also factors. The set of all factors of w of length $1 \leq i \leq |w|$ is denoted by: $Fact_i(w)$.

For $w = w_1 \cdots w_n$, where all the w_i 's belong to A , and for $i \leq j \leq |w|$, the notation $w[i \dots j]$ refers to the factor of the word w from position i to position j which is the factor $f = w_i w_{i+1} \cdots w_j$. We let $Pref_k(w) = w[1 \dots k]$ represents the *prefix* of length k of w . For all $1 \leq k \leq |w|$, $w[k]$ represents the k^{th} element of the word w .

We let $\tilde{w} = w_n \cdots w_2 w_1$ represents the **reversal** of the word w and we say that w is a **palindrome** if $\tilde{w} = w$. We remark that w is a palindrome if and only if $w_i = w_{n-i+1}$ for all $i \in \{1, \dots, |w|\}$. The set of all the palindromes over the alphabet A is denoted by: $Pal(A^*) = \{w \in A^*; \tilde{w} = w\}$.

Example 4 Let $w = 1001010010$ be a word in $\{0, 1\}^*$.

1. The reversal of w is $\tilde{w} = 0100101001$.
2. The factor $w' = w[2 \cdots 8] = 0010100 \in Pal(\{0, 1\}^*)$.

Two words w and w' are **conjugate** of order k , and denoted $w \equiv_k w'$ if and only if there exists u, v such that $|u| = k$ with $w = uv$ and $w' = vu$. When the exact value of k is not relevant, we simply write $w \equiv w'$.

Example 5 For the two words $w = 001100101$ and $w' = 100101001$ in $\{0, 1\}^*$, we have $w \equiv_3 w'$.

In addition to that, the circular permutation of a word w of length n at position i , denoted $\sigma_i(w)$, is defined by $\sigma_i(w_1 \cdots w_n) = w_{i+1} \cdots w_n w_1 \cdots w_i$ for all $0 \leq i \leq n - 1$.

Given a word $w = w_1 w'$ where w_1 is a letter, we denote by $w_1^{-1} w = w'$ the removal of the letter w_1 at the beginning of w . If w ends with letter w_1 , then the notation $w w_1^{-1}$ is defined accordingly.

The alphabet A can be ordered since we can define many order relations on words and we write $(A, <)$. In this thesis, we are interested only in the total **lexicographic order** or simply the *dictionary order* that is a total order and denoted: $<_\ell$

Definition 1.1 Let $w, w_1 \in A^*$, we say that:

$$w <_{\ell} w_1 \text{ if } \begin{cases} Pref_{|w|}(w_1) = w; \text{ or} \\ w = uav \text{ and } w_1 = ubv_1 \text{ where } u, v, v_1 \in A^* \text{ and } a <_{\ell} b, . \end{cases}$$

Example 6 We consider the following words: $w = 001000$, $w' = 001$ and $w'' = 001101$. We remark that $w <_{\ell} w''$, $w' <_{\ell} w''$ and $w' <_{\ell} w$.

In combinatorics on words, we can find many families of words over a finite alphabet. We have the family of Sturmian words that generates infinite words. The family of Christoffel words that is restricted to the family formed by finite Sturmian words. We will be interested also in a third family called family of Lyndon words defined in the next section.

1.2 Lyndon words

In 1954 and using the lexicographic order, Roger Lyndon [34] investigated the family of Lyndon words after being introduced by Anatoly Shirshov [45] in 1953 with a different name. The proper definition of this family of words is stated as follows:

Definition 1.2 A word $w \in A^*$ is a Lyndon word if for all $u, v \in A^+$ such that $w = uv$, we have $w <_{\ell} vu$.

In other words, we can say that the Lyndon word is the smallest word among all its conjugates. Hence, since $\sigma_i(w)$ is the circular permutation of w at position i which gives the conjugate of w of order i , hence we can write $w <_{\ell} \sigma_i(w)$, $\forall i \in \{1 \dots |w|\}$. It is equivalent that for each such factorization $w <_{\ell} v$.

Based on this definition, we can deduce one of the properties of this family of words. In fact, we have that *Lyndon* words are primitive words otherwise, we can have an equality between at least one of the conjugates of w . Therefore, if w is a Lyndon word, ww is not Lyndon.

Example 7 Let $w = 0110$, we have $X = \{\sigma_i(w); \forall 0 \leq i \leq 3\}$, hence we can write $X = \{0011, 0110, 1100, 1001\}$. The smallest element of X in lexicographic order is $w' = 0011$. Therefore, w is not a Lyndon word.

Over the alphabet $A = \{0, 1\}$, we represent the first 9 Lyndon words where the trivial words are the letters of the alphabet.

$$0, 1, 01, 001, 011, 0011, 0111, 0001, 0001, \dots$$

We say that s is a proper suffix of w if we can write $w = w's$ where $s \neq \epsilon$ and $w \neq w'$.

We define now two types of factorization for the Lyndon words, the **right** and the **left** factorizations. We mention before, that two words u and v are elements of the factorization of w if w is obtained by the concatenation of these two words. We represent the factorization by $w = uv = (u, v)$.

Proposition 1.3 *Let w be a Lyndon word, if w can be written as $w = (u, v)$ with v its lexicographically least proper suffix, then u, v are Lyndon words and $u <_l v$.*

This is the **right factorization**, while if u is the longest prefix then we get the **left factorization**.

In some cases these two factorizations coincide as we will see in Example 8 and in this case the word is called a **regular** word. This concept was used by Melançon in 1999 [35], then presented by (Berstel et al.2009) in [5] with a different notation as “*balanced₂*” and finally used by (Reutenauer 2016) [43] as “**equilibrated**”.

Example 8 *Let $w_1 = xyxxzxy$, $w_2 = xxyy$, and $w_3 = xyxxyxy$ be three words over the alphabet $A = \{x, y, z\}$. Table 1.1 shows the right and left factorization of each word. We remark that the left and right factorization of w_3 are the same,*

Word	right factorization	left factorization
w_1	$(xy, xxzxy)$	$(xyxxz, xy)$
w_2	(x, xyy)	(xxy, y)
w_3	$(xy, xxyxy)$	$(xy, xxyxy)$

Table 1.1: The right and left factorization of $w_1 = xyxxzxy$, $w_2 = xxyy$, and $w_3 = xyxxyxy$.

hence w_3 is a regular word.

The right factorization is simply called the **standard factorization**. We give now another kind of factorization that was defined in 1958 by Chen, Fox and Lyndon [13].

Definition 1.4 *Every non empty word w admits a unique factorization as a lexicographically decreasing sequence of Lyndon words: $w = l_1^{n_1} l_2^{n_2} \cdots l_k^{n_k}$, s.t $l_1 >_l l_2 >_l \cdots >_l l_k$ where $n_i \geq 1$ and l_i are Lyndon words.*

This factorization is called **Lyndon factorization**. In 1980 Duval proved that this factorization can be computed in a linear time [22].

The word	The Lyndon factorization
w_1	$(1)(001011)(0010101)(0)$
w_2	$(001)(001)(0) = (001)^2(0)$
w_3	$(00101011)(001)(0)$

Table 1.2: The Lyndon factorization of $w_1 = 100101100101010$, $w_2 = 0010010$ and $w_3 = 001010110010$.

Example 9 In Table 1.2, we find the Lyndon factorization of the following three words:

$w_1 = 100101100101010$, $w_2 = 0010010$ and $w_3 = 001010110010$.

Lyndon words give rise to a basis of some algebra. In particular, those factorizations were used as tools to solve problems in free groups and free Lie algebra [42].

1.3 Standard Sturmian Words

In this section, we study another family of words. We introduce the family of **standard Sturmian** words as the family of one sided infinite words over a binary alphabet that are obtained as a discretization of a line with irrational slope starting from the origin. An old game of English billiards on a square table can clearly represent the situation. In fact, the pounded ball follows a trajectory and hits the billiard board in such way that the vertical intersections are denoted by 0 and the horizontal ones are denoted 1. We represent this trajectory in a square lattice by considering the linear trajectory of irrational slope that cuts the squared lattice and we label the horizontal intersection by 1 and the vertical one by 0. The sequence obtained forms exactly the standard Sturmian word.

By considering the finite case of standard Sturmian words, the slope of the line segments turns into a rational number and we get a new family called family of **Christoffel words** that will be the fundamental object of this thesis. All the properties and characteristics of this family will be given in detail in the next section.

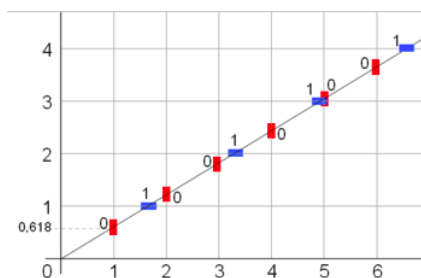


Figure 1.1: Beginning of a standard Sturmian word with an irrational slope.

One of the most important characterizations of the standard Sturmian words is the order of balancedness that is defined as follows:

Definition 1.5 [51] *A word w is k -balanced if, for any two factors u, v of w , of same length, and any letter \mathbf{a} of the alphabet A ,*

$$\text{If } \forall 1 \leq i \leq |w|, u, v \in \text{Fact}_i(w); |u| = |v| \implies ||u|_{\mathbf{a}} - |v|_{\mathbf{a}}| \leq k.$$

By convention, a balanced word refers to a 1-balanced word, see [40]. One of the properties of standard Sturmian words is that are balanced words. Further more in this chapter, we will expand and explicit the balancedness property.

While studying the structure of these words, we can not avoid to mention their periodicity defined as follows:

Definition 1.6 *A positive integer p is a **period** of a finite word w ; $1 \leq p \leq |w|$ if $w[i] = w[i + p]$; for all $1 \leq i \leq |w| - p$.*

A finite word can have more than one period and at the same time, we can find some words with no periods.

Example 10 *Let $w = 0001$, with Definition 1.6, we can note that for all $1 \leq i \leq 4$, the word w has no period.*

The most important result concerning the periodicity is the theorem Fine and Wilf (1965) [23] that describes the structure of words with more than one period.

Theorem 1.7 *If w has two periods p and q such that $|w| \geq p + q - \text{gcd}(p, q)$ then the $\text{gcd}(p, q)$ is also a period of w .*

Example 11 *Let $w = 0100100100100$ be a word of length $n = |w| = 13$. As we can see, the word w has a first period $p = 6$ since $w[1 \dots 7] = 0100100 = w[7 \dots 13]$. We also remark that the second period of w is $q = 9$ where $w[1 \dots 4] = 0100 = w[10 \dots 13]$. Hence this word has two periods $p = 6$ and $q = 9$ and since $n \geq 6 + 9 - \text{gcd}(6, 9) = 12$, then by the theorem Fine and Wilf, we have that $3 = \text{gcd}(6, 9)$ is also a period as we can see: $w[1 \dots 10] = w[4 \dots 13] = 0100100100$.*

This theorem has an optimal bound related to the case when the two periods p and q are coprime. In this case, the maximal length of a non constant word is $p + q - 2$. Such words are called *extremal Fine and Wilf words*. This is the second characterization of Sturmian words, since in 1994, de Luca and Mignosi in [17] showed that the set of all factors of extremal Fine and Wilf words corresponds to the set of factors of standard Sturmian words.

1.4 Christoffel words

After defining the previous two families of words, we study in detail and we give all the properties of this particular family of words over a binary alphabet. Christoffel words [14] have many equivalent definitions. In 1771, Jean Bernoulli was the first to give the definition of Christoffel words in the discrete plane and in 1990 Jean Berstel gave them this name with respect to Elwin. B. Christoffel (1829 – 1900). In a geometric view, we consider the Christoffel words to be the discretization of a line segment of rational slope. This was well explored in the papers (Osborne and Zieschang, 1981 [37]; Dulucq and Gouyou-Beauchamps, 1990 [21]; Borel and Laubie, 1993 [7]; Berstel and al., 2009 [5]). For the rest of this thesis, we restrict the work to the binary alphabet $A = \{0, 1\}$, unless mentioned elsewhere.

We start by defining the Christoffel path which is defined on a discrete lattice, i.e. it is a sequence of unitary steps joining points of integer lattice.

We recall that if two integers a and b in the set of non-negative integers \mathbb{N} have the property that their difference $a - b$ is integrally divisible by a number n (i.e., $(a - b)/n$ is an integer), then a and b are said to be "*congruent modulo n .*" and we denote $a \equiv b \pmod{n}$. Two numbers a and b are **coprime** if the greatest common divisor between them is 1, and we denote it: $a \perp b$.

Let $a \perp b$, the *Lower* Christoffel path of slope a/b is the path joining the origin $O(0, 0)$ to the point (b, a) and respecting the following characteristics:

- (i) it is the nearest path below the line segment joining these two points;
- (ii) there are no points of $\mathbb{Z} \times \mathbb{Z}$ between the path and line segment.

Analogously, the *Upper* Christoffel path is the path that lies above the line segment. By convention, the Christoffel path is exactly the *Lower* Christoffel path.

Example 12 Consider the line segment joining the origin $O(0, 0)$ to the point $(8, 5)$. We have $a = 5$, $b = 8$ and $n = a + b = 13$. The *Lower* and *upper* Christoffel paths of slope $5/8$ are represented in the Figure 1.2:

We remark that the Upper Christoffel path of slope a/b is the mirror image of the Lower Christoffel path obtained by a rotation of an angle of 180° .

By assigning to each horizontal step the letter 0 and to each vertical step the letter 1 of the binary alphabet $A = \{0, 1\}$, we get the *Christoffel word* of slope $\frac{a}{b}$ denoted by: $C(\frac{a}{b})$, where the fraction $\frac{a}{b}$ is exactly: $\frac{|w|_1}{|w|_0}$ (see Figure 1.3).

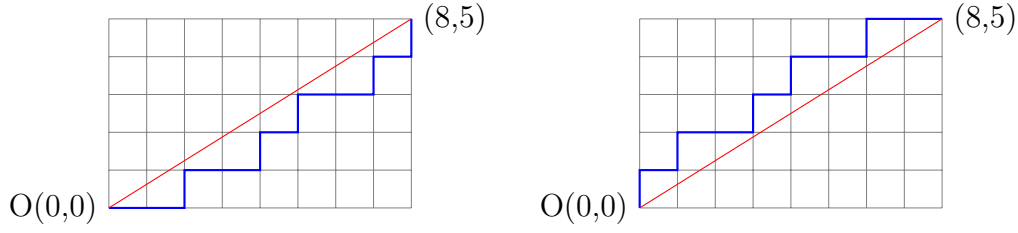


Figure 1.2: The Lower and Upper Christoffel paths of slope 5/8 respectively.

We define the morphism $\rho : A^* \rightarrow \mathbb{Q} \cup \{\infty\}$ by:

$$\rho(\epsilon) = 1 \text{ and } \rho(w) = \frac{|w|_1}{|w|_0}, \forall w \neq \epsilon \in A^*;$$

where $\frac{1}{0} = \infty$. This morphism determines the slope for each given word in A^* . In particular:

1. For $a = 0$, we get the Christoffel word of slope 0 represented by $w = C(\frac{0}{1}) = 0$.
2. For $b = 0$, we have the Christoffel word of slope ∞ represented by $w = C(\frac{1}{0}) = 1$.

Example 13 Let $(a, b) = (3, 5)$, the Christoffel word of slope 3/5 is given by: $C(\frac{3}{5}) = 00100101$ and represented in Figure 1.3:

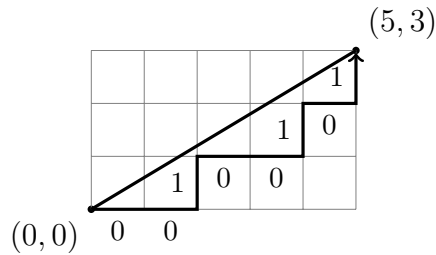


Figure 1.3: Illustration of the geometrical definition of Christoffel words. The Christoffel path goes from $(0,0)$ to $(5,3)$ and $C(\frac{3}{5}) = 00100101$.

Let $A = \{0, 1, \dots, k\}$ be an alphabet with integers, we let E be the automorphism defined as follows: $\forall i \in A$;

$$E : A \rightarrow A$$

$$i \rightarrow k - i$$

We have $E(i) + E(k - i) = k$. In particular, as we are dealing with the binary alphabet $A = \{0, 1\}$, we have: $E(0) = 1$ and $E(1) = 0$.

Example 14 Let $u = 00100110$ and $v = 11011001$, we have $v = E(u)$.

Remark 1.8 With the automorphism E , we remark that if w is a Lower Christoffel word of slope a/b then $E(w)$ is an Upper Christoffel word of slope b/a .

Example 15 For the Lower Christoffel word of slope $3/5$, we get $w = 00100101$. By evaluating $E(w)$, we get: $E(w) = 11011010$ which is an Upper Christoffel word of slope $5/3$ as shown in Figure 1.4.

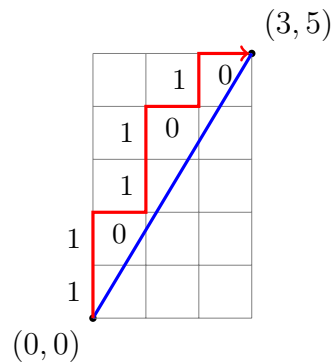


Figure 1.4: Illustration of the geometrical definition of Upper Christoffel words. The Christoffel path goes from $(0,0)$ to $(3,5)$ and $C(\frac{5}{3}) = 11011010$.

In 1997, Berstel in his paper [4], gave 14 different properties for the Christoffel words. Here, we mention four of them that are useful for the results obtained in this thesis.

Property 1.9 *Christoffel words are primitive words.*

In fact, if w is a Christoffel word of slope a/b then $|w|_1 = a$ and $|w|_0 = b$ where $a \perp b$. This implies that w is not a proper power of another word which means that w is a primitive word.

Property 1.10 [5] *Let $w = C(\frac{a}{b})$ be the Christoffel word of slope a/b , we write $w = 0w'1$, where w' is a palindrome.*

We name w' the central part of w . Note that the lower and upper Christoffel words have the same central part.

For example: The word $w = C(\frac{3}{5}) = 0\boxed{010010}1$ is a Christoffel word, where the central part 010010 is a palindrome as shown in the Figure 1.3.

In the previous section, we defined the Lyndon word to be the smallest between its conjugates. This property may be equivalently stated for Christoffel words and hence we have the third characterization of Christoffel words in terms of Lyndon words [4].

Theorem 1.11 [35] *A word w is a Christoffel word if and only if it is a 1-balanced Lyndon Word.*

We remark that the conjugate of Christoffel words are all distinct.

Example 16 *Let $w = C(\frac{2}{3}) = 00101$ be the Christoffel word of slope $2/3$. In Table 1.3, we show all the conjugates of w placed in an increasing lexicographic order:*

00101
01001
01010
10010
10100

Table 1.3: The lexicographic order of all the conjugates of $w = C(\frac{2}{3}) = 00101$.

We remark that the first word in the lexicographic order is the Christoffel word of slope $2/3$, while the last word in the lexicographic order is exactly the Upper Christoffel word of slope $2/3$. Hence, we can conclude that Lower and Upper Christoffel words are conjugate as we can see in Example 16 and this result was proved in [17] by De Luca and Migonsi in 1994.

1.4.1 Standard factorization

One of the most important characterizations of a Christoffel word is that any Christoffel word can be written as the concatenation of two Christoffel words in a unique way. This concept was introduced by Borel and Laubie in 1993 [7] and they called it the **Standard factorization** defined on proper Christoffel words as follows:

Theorem 1.12 (Borel-Laubie 1993) [7] *A proper Christoffel word w has a unique standard factorization $w = (u, v)$, where u and v are both Christoffel words.*

For this theorem, we are able to find several proof in different books. I introduce the proof given by Hugh Thomas that was mentioned in [5].

Proof. Let w be a Christoffel word of the line segment $[OF]$ of slope a/b . We prove the existence and uniqueness of this factorization as follow:

1. The existence:

Let C be the closest point on the Christoffel path to the line segment $[OF]$. The triangle OCF has no interior integer points, same for the two paths determined by the line segments $[OC]$ and $[CF]$ as we can see in the Figure 1.5. The consequence of this statement is that the coordinates of C are coprime and the two paths obtained are Christoffel ones. Hence $w = (u, v)$, where u is the Christoffel word corresponding to the line segment $[OC]$ and v is the Christoffel word corresponding to the line segment $[CF]$.

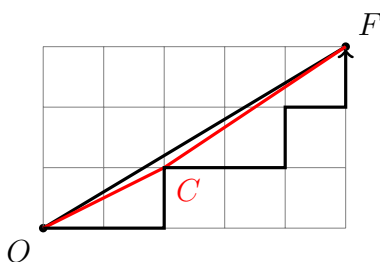


Figure 1.5: The standard factorization at the point C .

2. The uniqueness:

Let P be another point of the Christoffel path where w can be factorized as $w = (u', v')$. We have that the two triangles OCF and OPF have no interior integer points and h , the distance of the point C to $[OF]$, is different than h' , the distance of the point P to $[OF]$ as we can see in Figure 1.6. We recall the **PICK's theorem** stated as follows:

Let ABC be a triangle, we have: $area_{ABC} = i + \frac{1}{2}b - 1$, where i is the number of the interior points of ABC and b represents the number of the lattice point on this triangle.

By applying this theorem to the triangles OCF and OPF , we get for both triangles the same area d that corresponds to: $d = 0 + \frac{3}{2} - 1 = \frac{1}{2}$. Since the two triangles have two different heights, this leads us to a contradiction.

□

From this theorem we can conclude that the standard factorization is always obtained at the point C of the Christoffel path which is the closest point with respect to the line segment.

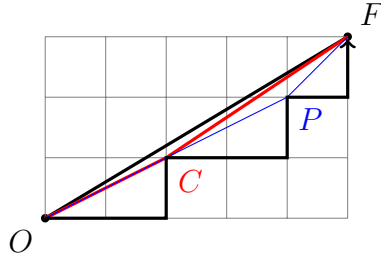


Figure 1.6: The two triangles OPA and OCA .

Example 17 Let w be the Christoffel word of slope $5/8$. We let P be the closest point to the line segment on the Christoffel path. We have $w = 0010010100101$ and the standard factorization is given by: $w = (w_1, w_2) = (00100101, 00101)$, where w_1 and w_2 are Christoffel words as represented in Figure 1.7.

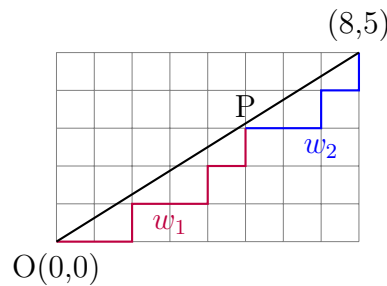


Figure 1.7: The standard factorization of the Christoffel word of slope $5/8$.

A direct application of the last theorem and the Property 1.10 is the following characteristic by (Chuan, 1997) [15] and called the Palindromic factorization. We recall that $Pal(A^*)$ is the set of all the palindromes over the alphabet A , and we have:

Property 1.13 Let w be a Christoffel word with (u, v) its standard factorization. The palindromic factorization is given uniquely by writing $w = p_1 p_2$, where p_1 and p_2 belongs to $Pal(A^*)$.

The proof can be obtained starting from the standard factorization of w . In fact, we have $w = (u, v)$ where u and v are Christoffel words. By applying the Property 1.10 to each of the Christoffel words w, u and v , we obtain the results as described in the Figure 1.8.

w					
u			v		
0	u_1	1	0	v_1	1
0	w_1				1
0	v_1	0	1	u_1	1
p_1			p_2		

Figure 1.8: The words, w , u and v are Christoffel words with palindromic central parts w_1, u_1, v_1 resp. The property of palindromes gives the Palindromic factorization of w denoted by $w = (p_1, p_2)$.

Example 18 Let $w = C(\frac{5}{8}) = 0010010100101$, the standard factorization as seen in Example 17 is: $w = (00100101, 00101)$ while the palindromic factorization is the following: $w = (00100, 10100101)$.

We can realize that since the standard factorization uses the closest point of the Christoffel path to the line segment, then by construction, the Palindromic factorization is exactly obtained using the furthest point of the Christoffel path. Those two points are very important points in this thesis. We enlighten on the importance of the first point in the following section, while the furthest point will be considered in Chapter 5 and 7 of this thesis.

1.4.2 Christoffel tree

At the beginning of this section, and in order to be able to construct the Christoffel tree, we introduce four morphisms called Christoffel morphisms that are in particular endomorphisms of the free monoid $\{0, 1\}^*$.

Those endomorphisms assign each Christoffel word onto a conjugate of a Christoffel word. We recall that an endomorphism f of A^* is defined by a couple (w_1, w_2) such that $f(0) = w_1$ and $f(1) = w_2$. So for the word $w = a_0 a_1 \cdots a_r \in A^*$, we have $f(w) = f(a_0) f(a_1) \cdots f(a_r)$. The four endomorphisms are defined as follows:

$$G = (0, 01); \quad D = (10, 1); \quad \tilde{G} = (0, 10) \text{ and } \tilde{D} = (01, 1)$$

We remark from the construction of these endomorphisms and the fact that every endomorphism maps conjugate words to conjugate words, the composition of these endomorphisms give rise to one of them. Therefore, these morphisms are called Christoffel morphisms, which means the image of a Christoffel word is a Christoffel word.

In particular if we take the endomorphism G , it assigns the Christoffel word of slope $\frac{a}{b}$ into the Christoffel word of slope $\frac{a}{a+b}$. While the endomorphism \tilde{D} , it assigns the Christoffel word of slope $\frac{a}{b}$ into the Christoffel word of slope $\frac{a+b}{b}$ as we can see in the Example 19.

Example 19 Let $w = C(\frac{3}{5}) = 00100101$, by calculating $G(w)$ and $\tilde{D}(w)$, we get:

$$\begin{aligned} G(w) &= G(00100101) = G(0)G(0)G(1)G(0)G(0)G(1)G(0)G(1) \\ &= 00010001001 \\ &= C(\frac{3}{8}). \end{aligned}$$

$$\begin{aligned} \tilde{D}(w) &= \tilde{D}(00100101) = \tilde{D}(0)\tilde{D}(0)\tilde{D}(1)\tilde{D}(0)\tilde{D}(0)\tilde{D}(1)\tilde{D}(0)\tilde{D}(1) \\ &= 0101101011011 \\ &= C(\frac{8}{5}). \end{aligned}$$

From those four endomorphisms, we choose the endomorphisms G and \tilde{D} to construct the Christoffel tree.

The **Christoffel tree** is an infinite binary tree whose elements are Christoffel words. The root of this tree is the Christoffel word of slope 1/1 represented by 01. The elements are written in their standard factorization form, hence each element has the form of a couple starting by $(0, 1)$ and respecting the rule given in Figure 1.9.

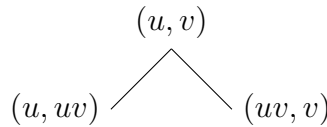


Figure 1.9: The rule followed to construct the Christoffel tree.

We remark that the root of the tree is the Christoffel word of slope 1. Then each node (w_1, w_2) has two sons: $G(w_1, w_2)$ on the left and $\tilde{D}(w_1, w_2)$ on the right. In Figure 1.10, we illustrate the first levels of the Christoffel tree:

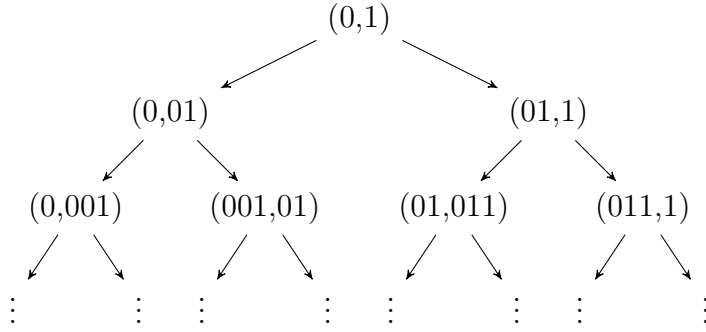


Figure 1.10: The first levels of the Christoffel tree.

1.5 Balance property

In Definition 1.5, the order of balancedness is obtained by studying the balancedness of the word over all the letters of the alphabet. Since we are dealing with the binary alphabet $A = \{0, 1\}$, it is enough to study the balancedness over one letter to deduce the order of balancedness. In this section, we introduce a matrix called the **Balance matrix** that studies the order of balancedness for any binary word. In fact, in this thesis, we choose the letter 1 to detect this order and hence we have: A word $w \in \{0, 1\}^*$ is k -balanced if and only if for every factors u, v of w , we get:

$$|u| = |v| \implies \left| |u|_1 - |v|_1 \right| \leq k.$$

We recall that for any matrix M , the i^{th} row of M is denoted: $M[i]$. By abuse of notation, a row $M[i]$ is seen as a word where each entry of the matrix is a letter.

The notation w^ω corresponds to the **circular word** w or simply saying $w \cdots w$. The **Language** of w index i is the set of all the factors of length i of the word w . Therefore, we can write: $L_i(w) = \{w' | w' \in \text{Fact}_i(w)\}$.

Example 20 For the word $w = 00101$, the circular factors are:

i	$L_i(w^\omega)$
0	ϵ
1	0, 1
2	00, 01, 10
3	001, 010, 101, 100
4	0010, 0101, 1010, 0100, 1001
\vdots	\vdots

Table 1.4: The circular factors of the word: $w = 00101$.

By [8], we have that a word w of length n is primitive if and only if the word w^ω has at least $k + 1$ factors of length k for all $k \in \{0, \dots, n - 1\}$.

1.5.1 Balance matrix

After defining the balancedness, we introduce in this section, a new matrix called the *Balance matrix* that gives the order of balancedness for a binary word. In fact, it is an explicit technique that explains the approach and the process followed in order to determine the balancedness of a binary word.

To get the Balance matrix, we have to define in a first step the matrix S whose entries $S[i, j]$ count the number of occurrences of letter one in all the factors of the circular binary word. So for $w^\omega = ww \dots w \dots$, we let:

$$s_w(\ell, p) := w^\omega[p \dots p + \ell - 1]$$

be the factor of length ℓ starting at position p in w where w is seen as a circular word with $1 \leq \ell, p \leq n$ and $n = |w|$. This is a refinement of the circular permutation $\sigma_i(w)$, since $s_w(\ell, p)$ gives the exact factor at position p and of length ℓ .

With this notation, we restrict the *language* of w^ω to words of length ℓ and we denote: $L_\ell(w^\omega) = \{s_w(\ell, p) \mid 1 \leq p \leq |w|\}$.

Using those factors, we define the matrix S_w where each element at position (i, j) is the number of occurrences of the letter 1 in the factor of length i and starting at position j . Hence, we can define the $n \times n$ matrix S_w as follows:

$$S_w[i, j] = |s_w(i, j)|_1.$$

Example 21 Let $w = C(\frac{3}{5}) = 00100\mathbf{101}$, if we take all the factors of length 3 and we count the number of occurrences of the letter 1, we get the following row: $S_w[3] = 11111211$. The element 2 corresponds to the factor 101 at position 6; i.e $S_w[3, 6] = |s_w[3, 6]|_1 = |101|_1 = 2$. If we repeat the same method for all the factors and at all the positions, we get:

$$S_w = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & \mathbf{2} & 1 & 1 \\ 1 & 1 & 2 & 1 & 2 & 2 & 1 & 2 \\ 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 3 & 2 & 2 & 3 & 2 & 2 \\ 2 & 3 & 3 & 2 & 3 & 3 & 2 & 3 \\ 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \end{pmatrix}$$

Remark 1.14 We can see that by considering a sliding window of length i with $1 \leq i \leq n$ and starting from position 1 to position n , the number of occurrences of the letter 1 in this window determines the entries of the row $S_w[i]$.

From Theorem 1.11, we know that a Christoffel word is a 1-balanced Lyndon word. The following technical lemma states that the prefix of a Christoffel word always realizes the minimum number of occurrences of letter 1 among all the factors of same length.

Lemma 1.15 For any length i and any Christoffel word w ,

$$|s_w(i, 1)|_1 = \min_j (|s_w(i, j)|_1).$$

Proof. By contradiction, suppose that a Christoffel word w admits a factor u of length l that is not a prefix and that $|u|_1 < |w[1..l]|_1$. Moreover, suppose that u is the shortest factor of w to have this property. Consequently, u ends with letter 0 while the prefix $w[1..l]$ ends with letter 1. By Theorem 1.11, w is Lyndon and 1-balanced. Let s be a suffix of w that admits the factor u as a prefix. Moreover, let p be the longest common prefix to u and w . Since s is lexicographically smaller than w , it must be that $p0$ is a prefix of w while $p1$ is a prefix of u . From the above observations, we conclude that there exist three words u' , w' , w'' such that $|u'| = |w'|$, $u = p1u'0$ and $w = p0w'1w''$.

$$\begin{array}{l} w = \begin{array}{|c|c|c|c|c|} \hline p & 0 & w' & 1 & w'' \\ \hline \end{array} \\ u = \begin{array}{|c|c|c|c|} \hline p & 1 & u' & 0 \\ \hline \end{array} \end{array}$$

Finally, $|u|_1 < |w[1..l]|_1$ implies that $|u'|_1 < |w'|_1$. This is impossible since w is a 1-balanced word having $w'1$ and $u'0$ as factors, both have same length and $|w'1|_1 \geq |u'0|_1 + 2$. Contradiction. \square

In the following, we introduce the *balance matrix* B_w , that is obtained from S_w by subtracting the minimum value on each row.

Definition 1.16 Let w be a binary word of length n , the balance matrix of w is:

$$B_w[i, j] = S_w[i, j] - \min(S_w[i]), \quad \forall 1 \leq i, j \leq n$$

For the rest of this thesis, we denote the order of balancedness of w by δ and we give in Definition 1.17 an equivalent definition for this order:

Definition 1.17 We say that w is δ -balanced if:

$$\delta = \max_i (\max(S_w[i]) - \min(S_w[i])) = \max(B_w); \quad \forall 1 \leq i \leq |w|.$$

Example 22 For the same Christoffel word w of slope $3/5$ mentioned in Example 21, we calculate its balanced matrix $B_{\frac{3}{5}}$ and we get:

$$S_w = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 & 2 & 2 & 1 & 2 \\ 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 3 & 2 & 2 & 3 & 2 & 2 \\ 2 & 3 & 3 & 2 & 3 & 3 & 2 & 3 \\ 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \end{pmatrix} \Rightarrow B_w = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Since the order of balancedness δ of a binary word w is the maximal difference of occurrences of the letter 1 in all the factors of same length from Definition 1.29. This leads to realize that the maximal value obtained in the balance matrix B_w represents exactly δ . Hence, we obviously can say that w is $(\max B_w)$ -balanced.

Remark 1.18 We can mention some remarks concerning the balanced matrix:

1. By construction, if $w \notin 1^*$ then the first row of B_w is equal to w .
2. If $k < |w|$ is a period of w^ω , again by construction we have that the row $B_w[k]$ contains only zeros. The converse is also true, an integer k such that the row $B_w[k]$ contains only zeros is a period of w^ω .

Consequently, a row of zeros is called a **period row**.

Property 1.19 Let w be a binary word, the i^{th} row of B_w , where $i = n = |w|$ is always a period of w^ω , therefore $B_w[n]$ is a period row.

Proof. Let w be a binary word with $n = |w|$. By calculating $S_w[n]$, we are looking for all the factors of w^ω of length n that contain exactly the same number of occurrences of letter 1 as the word itself. Suppose that $|w|_1 = a$, then $S_w[n] = a^n$. By applying Definition 1.16, we get $B_w[n] = 0^n$ and we have a period row. \square

From this property, we can conclude that the matrix B_w is a $(n-1) \times n$ matrix since the n -th row does not add any information.

Example 23 From Example 21, the n^{th} row of $w = 00100101$ is $S_w[8] = 33333333$. By applying Definition 1.16, we have to subtract by $3 = \min(S_w[8])$. Hence we get a period row.

Remark 1.20 In order to enlighten the notations, when working with the Christoffel $C\left(\frac{a}{b}\right)$, the balance matrix $B_{C\left(\frac{a}{b}\right)}$ is simply denoted $B_{\frac{a}{b}}$.

Example 24 Let $\frac{a}{b} = \frac{5}{2}$ and $n = 7$, we have $w = C\left(\frac{5}{2}\right) = 0110111$. To construct the balance matrix B_w , we take each row of S_w and we subtract its elements by the minimal value of this row and we get the following:

$$S_w = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 2 & 1 \\ 2 & 2 & 2 & 2 & 3 & 2 & 2 \\ 2 & 3 & 3 & 3 & 3 & 3 & 3 \\ 3 & 4 & 4 & 3 & 4 & 4 & 3 \\ 4 & 5 & 4 & 4 & 5 & 4 & 4 \\ 5 & 5 & 5 & 5 & 5 & 5 & 5 \end{pmatrix}, \quad B_{\frac{5}{2}} = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

Example 25 Let $w = 0011$ be a word on the binary alphabet $A = \{0, 1\}$. The infinite word w^ω is a 2-balanced word and we have $\max(B_w) = 2$.

$$S_w = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 2 & 1 \\ 1 & 2 & 2 & 1 \\ 2 & 2 & 2 & 2 \end{pmatrix}, \quad B_w = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}.$$

Some trivial questions can be asked after defining the matrix B .

1. What happens if the alphabet is not binary?
2. How are we able to determine the order of balancedness of a word over a d -alphabet?

In fact, in order to reply to the second question, we assign for each letter its proper balance matrix and we let the order of balancedness to be the maximal value obtained between those matrices.

Definition 1.21 Let $A = \{a_1, a_2, \dots, a_d\}$ and w a circular word in A^* . We generalize the form of B_w and we write $B_w^{a_k}[i, j] = |s_w[i, j]_{a_k}| - \min_i(S_w[i])$, $\forall 1 \leq k \leq d$. We say that w is δ -balanced if:

$$\delta = \max_i[B_w^{a_k}], \forall 1 \leq k \leq d.$$

Example 26 Let $w = 01221$ be a word over the alphabet $A = \{0, 1, 2\}$. w^ω is 2-balanced and we have:

$$B_w^0 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \end{pmatrix}, \quad B_w^1 = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix}, \quad B_w^2 = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 1 & 2 & 2 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{pmatrix}.$$

and hence: $\delta = \max(B_w^0, B_w^1, B_w^2) = 2$.

Sage code During those years of research, all the results were tested and computed using the Sage Software based on Python.

The matrix B_w is obtained by using the following code, where s denotes a binary word:

```

1 class OccurrenceCounter :
2     def __init__( self , l , a ) :
3         self.nb_occ = [0]
4         for i in l :
5             if i == a :
6                 self.nb_occ.append( self.nb_occ[-1] + 1 )
7             else :
8                 self.nb_occ.append( self.nb_occ[-1] )
9
10        def nbOcc( self , i , j ) :
11            return self.nb_occ[j] - self.nb_occ[i]
12
13    def Balance_Matrix( s ) :
14        oc = OccurrenceCounter( s+s , 1 )
15        n = len(s)
16        B = [ [ oc.nbOcc(j, j+i+1) for j in range(n)] for i in range(n-1)
17              ]
18        for i in range(n-1) :
19            m = min( B[i] )
20            B[i] = map( lambda x : x-m, B[i] )
21        return B

```

Listing 1.1: Code of the Balance Matrix.

Example 27 In Example 24 and using the definition of the balance matrix, we have determined all the entries of the balance matrix B_w where $w = C(\frac{5}{2}) = 0110111$. Using the code 1.5.1, we get the same result by computing:

$$\text{Balance_Matrix}([0, 1, 1, 0, 1, 1, 1]) = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

1.5.2 Properties of the balance matrix

In this section, we present some properties of the matrix B_w , where w is a binary word of length $|w| = n$, that allow us to restrict the work on the upper half of the balance matrix since the lower half is deduced by symmetry. For the Definition 1.22, we need to recall that $w \equiv_k w'$ if and only if there exists u, v such that $|u| = k$ with $w = uv$ and $w' = vu$ in order to prove one of the properties of the balanced matrix given in Lemma 1.26.

Definition 1.22 *Two words w and w' are complementary if $w \equiv_k E(w')$ for some k . Moreover, a word is said to be auto-complementary if it is complementary to itself.*

Example 28 *For the word $w = 0110110010$, we have $E(w) = 1001001101$, hence w is an auto-complementary word since $w \equiv_5 E(w)$. While $w = 0000100$ and $w' = 0111111$ are such that $w \equiv_4 E(w')$, where $E(w') = 1000000$.*

Lemma 1.23 *Any binary word w with $n = |w|$ can be written as:*

$$\sigma_{j-1}(w) = s_w[k, j]s_w[n - k, k + j]; \forall 1 \leq j \leq n \text{ and } 1 \leq k \leq n - 1.$$

Proof. Let w be a binary word with n its length. By induction we have:

- $s_w[k, 1] = w[1 \cdots k]$ for $p = 1$ and $\ell = k$;
- $s_w[n - k, k + 1] = w[k + 1 \cdots n]$ for $p = k + 1$ and $\ell = n - k$.

Therefore, we can write: $\sigma_0(w) = w[1 \cdots n] = s_w[k, 1]s_w[n - k, k + 1] = w[1 \cdots k]w[k + 1 \cdots n]; \forall k \in \{1, \dots, n - 1\}$. Suppose this is true for σ_{j-1} , we prove it true for σ_j .

We have $\sigma_{j-1}(w) = s_w[k, j]s_w[n - k, k + j]$, and we know that:

$$\begin{aligned} \sigma_j(w) &= a_j^{-1} \sigma_{j-1}(w) a_j \\ &= a_j^{-1} s_w^a[k, j] s_w^a[n - k, k + j] a_j \\ &= a_j^{-1} a_j a_{j+1} \cdots a_{k+j-1} a_{k+j} \cdots a_{j-1} a_j \\ &= a_{j+1} \cdots a_{k+j-1} a_{k+j} \cdots a_{j-1} a_j \\ &= s_w[k, j + 1] s_w[n - k, k - j + 1] \end{aligned}$$

□

Remark 1.24 *The number of occurrences of each letters in a word is stable by conjugacy, so that for all $1 \leq i, j \leq n$,*

$$S_w[i, j] + S_w[n - i, (i + j) \bmod n] = |w|_1. \quad (1.1)$$

Example 29 Let $a/b = 3/8$, $n = 11$; the circular Christoffel word of slope $3/8$ can be written as:

$$\sigma_2\left(C\left(\frac{3}{8}\right)\right) = 010001 \cdot 00100 = w^\omega[3, 8] \cdot w^\omega[9, 13].$$

w	w
00010001001	00010001001
$w^\omega[3, 8]$	$w^\omega[9, 13]$

So that:

$$|w^\omega[3, 8]|_1 + |w^\omega[9, 13]|_1 = |s_w[6, 3]|_1 + |s_w[5, 9]|_1 = S_w[6, 3] + S_w[5, 9] = |w|_1 = 3 = a.$$

Property 1.25 The matrix S_w is such that for any $1 \leq i < n$, either:

- (a) $\max S_w[i] + \min S_w[n - i] = |w|_1$,
- (b) $\min S_w[i] + \max S_w[n - i] = |w|_1$.

In particular, if w is a 1-balanced word, we have:

- (1) $\min S_w[i] + \min S_w[n - i] = |w|_1 - 1$,
- (2) $\max S_w[i] + \max S_w[n - i] = |w|_1 + 1$,

Proof. For (a), given $i \in \{1, \dots, n - 1\}$, let j be such that $S_w[i, j]$ is maximal on the i -th row. By equation (1.1), $S_w[n - i, i + j \bmod n]$ is minimal on the $(n - i)$ -th row, so that $\max S_w[i] + \min S_w[n - i] = a$. Equation (b) is shown in a similar way.

For (1) and (2), since w is a balanced word, thus $\max(S_w[i]) \leq \min(S_w[i]) + 1$. On the other hand, $\max(S_w[i]) \neq \min(S_w[i])$ since it would imply that i is a period row in B_w but this is impossible since w^ω does not have a period smaller than n . \square

Example 30 Let $w = C(\frac{5}{2})$, we have:

$$S_w = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 2 & 1 \\ 2 & 2 & 2 & 2 & 3 & 2 & 2 \\ 2 & 3 & 3 & 3 & 3 & 3 & 3 \\ 3 & 4 & 4 & 3 & 4 & 4 & 3 \\ 4 & 5 & 4 & 4 & 5 & 4 & 4 \end{pmatrix}$$

We can see that:

1. $\min S_w[1] + \max S_w[6] = 0 + 5 = 5,$
2. $\min S_w[3] + \min S_w[4] = 2 + 2 = 4,$
3. $\max S_w[2] + \max S_w[5] = 2 + 4 = 6.$

Lemma 1.26 *The balance matrix is such that for any $1 \leq i < n,$*

$$B_w[i] \equiv_{n-i} E(B_w[n-i]).$$

Proof. From the definition of $S_w,$ we have that the elements of $S_w[i] \in \{1, \dots, a_k\},$ where $a_k \leq |w|_1.$ Therefore, to prove that $B_w[i] \equiv_{n-i} E(B_w[n-i]),$ it is sufficient to prove that $B_w[i, j] = E(B_w[n-i, i+j]),$ or $B_w[i, j] + B_w[n-i, i+j] = a_k.$ Two cases are to be studied:

- First case, if $S_w[i, j] = \min S_w[i] = 0, \forall i < \frac{n}{2}:$
 In this case, $S_w[n-i, (i+j) \bmod n] = |w|_1 - S_w[i, j] = |w|_1,$ and $\min S_w[n-i] = |w|_1 - \max S_w[i]$ from Property 1.25.
 By the definition of B_w and equation (1.1),

$$\begin{aligned} B_w[i, j] &= |w|_1 - S_w[n-i, (i+j) \bmod n] - \min(S_w[i]) = 0, \\ B_w[n-i, i+j] &= S_w[n-i, (i+j) \bmod n] - \min(S_w[n-i]), \\ &= |w|_1 - (|w|_1 - \max S_w[i]) = a_k. \end{aligned}$$

By adding both rows, we get:

$$B_w[i, j] + B_w[n-i, i+j] = a_k.$$

- Second case, if $S_w[i, j] = a_i > \min S_w[i] = 0, \forall i < \frac{n}{2}:$
 In this case, $S_w[n-i, (i+j) \bmod n] = |w|_1 - S_w[i, j] = |w|_1 - a_i,$ and $\min S_w[n-i] = |w|_1 - \max S_w[i]$ from Property 1.25.
 By the definition of B_w and equation (1.1),

$$\begin{aligned} B_w[i, j] &= |w|_1 - S_w[n-i, (i+j) \bmod n] - \min(S_w[i]) = a_i, \\ B_w[n-i, i+j] &= S_w[n-i, (i+j) \bmod n] - \min(S_w[n-i]), \\ &= (|w|_1 - a_i) - (|w|_1 - \max S_w[i]) = a_k - a_i. \end{aligned}$$

By adding both rows, we get:

$$B_w[i, j] + B_w[n-i, i+j] = a_k.$$

- Third case, if $i = \frac{n}{2}$:

In this case, $S_w[\frac{n}{2}, (\frac{n}{2} + j) \bmod n] + S_w[\frac{n}{2}, j] = |w|_1$, and $\min S_w[\frac{n}{2}] + \max S_w[\frac{n}{2}] = |w|_1$. The elements of $S_w \in \{a_1, \dots, a_k\}$, therefore the elements of $B_w \in \{0, \dots, a_k - a_1\}$. Hence to reach the result, we have to prove that $B_w[\frac{n}{2}, j] + B_w[\frac{n}{2}, (\frac{n}{2} + j) \bmod n] = a_k - a_1$.

By the definition of B_w and equation (1.1),

$$\begin{aligned} B_w[\frac{n}{2}, j] &= S_w[\frac{n}{2}, j] - \min S_w[\frac{n}{2}], \\ &= |w|_1 - S_w[\frac{n}{2}, (\frac{n}{2} + j) \bmod n] - |w|_1 + \max S_w[\frac{n}{2}], \end{aligned}$$

$$B_w[\frac{n}{2}, (\frac{n}{2} + j) \bmod n] = S_w[\frac{n}{2}, (\frac{n}{2} + j) \bmod n] - \min S_w[\frac{n}{2}].$$

By adding both rows, we get:

$$B_w[\frac{n}{2}, j] + B_w[\frac{n}{2}, (\frac{n}{2} + j) \bmod n] = \max S_w[\frac{n}{2}] - \min S_w[\frac{n}{2}] = a_k - a_1.$$

□

A direct consequence of this lemma is that if $B_{\frac{a}{b}}$ has a middle row, then this row is auto-complementary.

Corollary 1.27 *If n is an even number, $B_{\frac{a}{b}}[\frac{n}{2}]$ is an auto-complementary row.*

Proof. The matrix $B_{\frac{a}{b}}$ has $n - 1$ rows that are pairwise complementary. by Lemma 1.26, the i^{th} row is complementary to the $(n - i)^{\text{th}}$ row. In particular for $i = \frac{n}{2}$, we have:

$$B_{\frac{a}{b}}[\frac{n}{2}] \equiv_{\frac{n}{2}} E(B_{\frac{a}{b}}[\frac{n}{2}]).$$

□

Example 31 *Let $a/b = 3/7$ with $n = 10$, the balance matrix $B_{\frac{3}{7}}$ is given by:*

$$B_{\frac{3}{7}} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

Note that $B_{\frac{3}{7}}[5]$ is an auto-complementary row and we have:

$$\begin{aligned} B_{\frac{3}{7}}[5] &= \boxed{00110 \mid 11001} \\ &= E(\boxed{11001 \mid 11001}) \end{aligned}$$

From the previous lemmas, the lower half of B_w is deduced from the upper half. The balance matrix associated to the rational number $\frac{1}{b}$ has a particular form.

Remark 1.28 *In the case where $a = 1$, the balance matrix is such that: $B_{\frac{1}{b}}[i] = 0^{(n-i)}1^i$.*

Example 32 *Let $w = C(\frac{1}{4})$ with $n = 5$, the balance matrix $B_{\frac{1}{4}}$ is of the following form:*

$$B_{\frac{1}{4}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Property 1.29 *Let $w = 0^x1^y$ such that $x, y > 0$. For $i \leq x$ and $i \leq y$, we have: $B_w[i] = 0^{x-i+1}12 \dots (i-2)(i-1)i^{y-i+1}(i-1)(i-2) \dots 21$.*

Proof. Let $w = 0^x1^y$, for $i = 2$, $S_w[i, p] = |w^\omega[p, p+i-1]|_1$ for all $1 \leq p \leq x+y$, therefore $S_w[2] = 0^{x-1}12^{y-1}1 = 0^{x-2+1}12^{y-2+1}1$ and since $x-1 > 0$ then $\min(S_w[2]) = 0$ and $B_w[2] = S_w[2]$. Suppose this is true for $i \leq x$, hence $B_w[i] = 0^{x-i+1}12 \dots i^{y-i+1} \dots 21$ prove it true for $i+1$ such that $i+1 \leq x$. We have $S_w[i+1, p] = 0$ for all $1 \leq p \leq x-i$, for $p = x-i+1$, we have $S_w[i+1, p] = 1$ and for $x-i+1 \leq p \leq y-x-i$, we get $S_w[i+1, p] = i+1$. The values decrease until $p = x+y$ and we have $S_w[i+1, x+y] = 1$. Since $i+1 \leq x$, then $x-i \geq 1$ and $\min S_w[i+1] = 0$, therefore $B_w[i] = 0^{x-i+1}12 \dots i^{y-i+1} \dots 21$. \square

Example 33 *Let $w = 000011111 = 0^41^5$, we have $x = 4$ and $y = 5$. For $i = 3$ such that $i \leq x$ and $i \leq y$, we have from Property 1.29:*

$$\begin{aligned} B_w[3] &= 0^2123^321 \\ &= 001233321. \end{aligned}$$

The same result is obtained if we take a sliding window of length 3 and counts the occurrences of the letter 1. Since the minimal value obtained is 0 hence the values obtained form the entries of the row $B_w[3]$.

1.5.3 Christoffel words with a and b not coprime

In all the previous properties we had $a \perp b$. In this section, we present the form of the balance matrix $B_{\frac{a}{b}}$, where a and b are not coprime. We start by noticing that the Christoffel word of slope a/b where $a = a'd$, $b = b'd$ and $d = \gcd(a, b)$ is the concatenation of d Christoffel words of slope a'/b' and we denote: $C(\frac{a}{b}) = [C(\frac{a'}{b'})]^d$.

Example 34 Let $w = C(\frac{6}{10})$, it is the Christoffel word of slope $6/10$ that discretized the line segment of slope $\frac{6}{10} = \frac{3}{5}$. But the two line segments of slope $6/10$ and $3/5$ represent the same line, hence w is the repetition of $w' = C(\frac{3}{5})$ as shown in Figure 1.11.

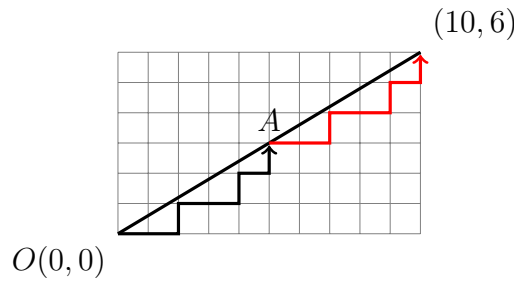


Figure 1.11: The Christoffel word of slope $\frac{6}{10}$.

We remark that w is obtained by repeating the word w' two times since $d = \gcd(6, 10) = 2$. Hence, we can say that w is the concatenation of w' and we write: $w = (w')^2$.

In the last part of this chapter, we introduce a new notation for matrices that is used for $B_{\frac{a}{b}}$ when a and b are not coprime. In fact, we denote $(M)_{\ell}^k$ the matrix having the block M repeated k times horizontally and ℓ times vertically.

Example 35 Let $M = \begin{pmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \end{pmatrix}$ and $M' = (M)_{\ell}^k = \begin{pmatrix} 0 & 1 & 2 & | & 0 & 1 & 2 \\ 3 & 4 & 5 & | & 3 & 4 & 5 \\ \hline 0 & 1 & 2 & | & 0 & 1 & 2 \\ 3 & 4 & 5 & | & 3 & 4 & 5 \\ \hline 0 & 1 & 2 & | & 0 & 1 & 2 \\ 3 & 4 & 5 & | & 3 & 4 & 5 \end{pmatrix}$.

The matrix M' is composed of the matrix M where M is repeated 3 times horizontally and 2 times vertically.

Property 1.30 Let $\frac{a}{b}$ be an unreduced fraction with $d = \gcd(a, b)$, the balance matrix $B_{\frac{a}{b}}$ contains exactly $d - 1$ zero rows.

Proof. Let $\frac{a}{b}$ be an unreduced fraction with $d = \gcd(a, b)$ and $n = a + b$, we introduce the three integers a', b', n' such that: $a = a'd, b = b'd$ and $n = n'd$. By construction, $C(\frac{a}{b}) = [C(\frac{a'}{b'})]^d$ and $B_{\frac{a'}{b'}}$ is composed of $n' - 1$ rows and n' columns where the n'^{th} row is the period row of this matrix.

Hence $B_{\frac{a}{b}}$ is obtained by the same matrix of $B_{\frac{a'}{b'}}$ repeated d times horizontally and vertically. Since for each balanced matrix we have the period row which is made of zeros and is omitted. Hence we deduce that $B_{\frac{a}{b}}$ is composed of exactly $d - 1$ zero rows since the last row of $B_{\frac{a'}{b'}}$ is removed. \square

Example 36 We let the fraction $\frac{4}{8}$ with $d = 4$, from Property 1.30, we get 4 repeated blocks all over the matrix $B_{\frac{4}{8}}$ and 3 zero rows and we write: $B_{\frac{4}{8}} = (B_{\frac{1}{2}})^4$.

$$B_{\frac{1}{2}} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \Rightarrow B_{\frac{4}{8}} = \begin{pmatrix} 0 & 0 & 1 & | & 0 & 0 & 1 & | & 0 & 0 & 1 & | & 0 & 0 & 1 \\ 0 & 1 & 1 & | & 0 & 1 & 1 & | & 0 & 1 & 1 & | & 0 & 1 & 1 \\ \color{red}{0} & \color{red}{0} & \color{red}{0} & | & \color{red}{0} & \color{red}{0} & \color{red}{0} & | & \color{red}{0} & \color{red}{0} & \color{red}{0} & | & \color{red}{0} & \color{red}{0} & \color{red}{0} \\ \hline 0 & 0 & 1 & | & 0 & 0 & 1 & | & 0 & 0 & 1 & | & 0 & 0 & 1 \\ 0 & 1 & 1 & | & 0 & 1 & 1 & | & 0 & 1 & 1 & | & 0 & 1 & 1 \\ \color{red}{0} & \color{red}{0} & \color{red}{0} & | & \color{red}{0} & \color{red}{0} & \color{red}{0} & | & \color{red}{0} & \color{red}{0} & \color{red}{0} & | & \color{red}{0} & \color{red}{0} & \color{red}{0} \\ \hline 0 & 0 & 1 & | & 0 & 0 & 1 & | & 0 & 0 & 1 & | & 0 & 0 & 1 \\ 0 & 1 & 1 & | & 0 & 1 & 1 & | & 0 & 1 & 1 & | & 0 & 1 & 1 \\ \color{red}{0} & \color{red}{0} & \color{red}{0} & | & \color{red}{0} & \color{red}{0} & \color{red}{0} & | & \color{red}{0} & \color{red}{0} & \color{red}{0} & | & \color{red}{0} & \color{red}{0} & \color{red}{0} \\ \hline 0 & 0 & 1 & | & 0 & 0 & 1 & | & 0 & 0 & 1 & | & 0 & 0 & 1 \\ 0 & 1 & 1 & | & 0 & 1 & 1 & | & 0 & 1 & 1 & | & 0 & 1 & 1 \end{pmatrix}.$$

Chapter 2

Algebraic view

In the previous chapter, we mentioned that this thesis is based essentially on Christoffel words and we gave the geometric and combinatoric definitions. Equivalently to all the definitions and properties mentioned before, we give in this chapter the algebraic definitions of Christoffel words and we show the link between those two points of view. This definition was initially given by Christoffel (1875) [14], and then mentioned in different articles like in [5, 7, 8].

2.1 Algebraic definition

A Christoffel word of slope a/b is a sequence of $a+b$ letters chosen from the binary alphabet $\{0, 1\}$. This selection of letters is not done randomly, in fact we are going to give three equivalent algebraic definitions that help us to determine the exact letter to be chosen.

The Christoffel word of slope a/b is denoted $w = C\left(\frac{a}{b}\right)$ and defined as follows:

Definition 2.1 [14] *Let $w = C\left(\frac{a}{b}\right)$ and $n = a + b$, for $0 \leq i \leq n$, let $r_i = ia \bmod n$. The word w is of length n and for $1 \leq i \leq n$, the i -th letter is given by:*

$$w[i] = \begin{cases} 0 & \text{if } r_{i-1} < r_i, \\ 1 & \text{otherwise.} \end{cases}$$

Example 37 *Let $w = C\left(\frac{5}{8}\right)$ of length $n = 13$. The sequence $(r_i)_{0 \leq i \leq n}$ is given as follows:*

$$r_0 = 0, \quad r_1 = 5, \quad r_2 = 10, \quad r_3 = 2, \quad r_4 = 7, \quad r_5 = 12, \quad r_6 = 4, \quad r_7 = 9, \quad r_8 = 1, \quad r_9 = 6, \quad r_{10} = 11, \quad r_{11} = 3, \quad r_{12} = 8, \quad r_{13} = 0.$$

Hence by Definition 2.1, we obtain:

$$w[1] = 0, \quad w[2] = 0, \quad w[3] = 1, \quad w[4] = 0, \quad w[5] = 0, \quad w[6] = 1, \quad w[7] =$$

0, $w[8] = 1$, $w[9] = 0$, $w[10] = 0$, $w[11] = 1$, $w[12] = 0$, $w[13] = 1$; and we write $w = 0010010100101$ as shown in Figure 2.1.

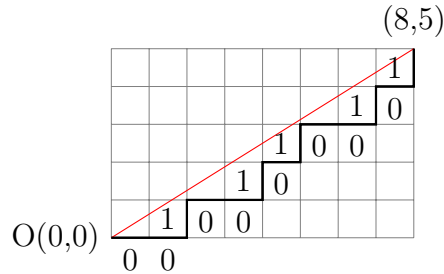


Figure 2.1: The Christoffel word of slope $5/8$.

In fact, the sequence $R = (r_i)_{0 \leq i \leq |w|}$, where w is the Christoffel word of slope a/b has a special geometric signification. Each integer point (k, l) on the Christoffel path is followed by:

1. Either the point $(k + 1, l)$ if we do a horizontal step, and hence the letter 0 is assigned to the word.
2. Or it is followed by $(k, l + 1)$ if we do a vertical step and in this case the letter 1 is chosen.

In 1875, Christoffel has introduced in [14] a relation that assigns to each point of the Christoffel path, and depending on its coordinates, a specific label. This relation is presented in Definition 2.2:

Definition 2.2 Let $w = C(\frac{a}{b})$ such that $a \perp b$ and $(b, a) \neq (0, 1)$. We assign to the integer point (k, l) on w the label: $\frac{ka-lb}{b}$.

This label is exactly the vertical distance obtained between the point of coordinates (k, l) and the line segment joining the origin to the point (b, a) .

Example 38 From Example 37, we have that the Christoffel word of slope $5/8$ is given by: $w = 0010010100101$ and its sequence $R = (r_i)_{0 \leq i \leq 13}$ is the following: $R = [0, 5, 10, 2, 7, 12, 4, 9, 1, 6, 11, 3, 8, 0]$. In Figure 2.2, we assign the labels to each integer point on this path. For the point of coordinates $(3, 1)$, the label assigned is exactly: $v = \frac{3 \cdot 5 - 8 \cdot 1}{8} = \frac{7}{8}$.

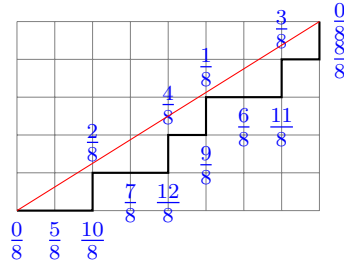


Figure 2.2: The labels assigned to each integer point of the Christoffel path of slope $5/8$.

Proposition 2.3 *Let a and b be two coprime numbers. If $\frac{s}{b}$ and $\frac{t}{b}$ are two labels assigned to two consecutive integer points on the Christoffel path of slope a/b , then $t = s + a \pmod{a + b}$. Also, t can take exactly one value from $0, 1, \dots, a + b - 1$ each time.*

From this proposition, we can get some conclusions:

1. The numerators of the labels are exactly the values of the sequence R .
2. The point P that refers to the closest point on the Christoffel path which is the point used to obtain the standard factorization is labeled by $\frac{1}{b}$. In other words, the point P is obtained for a certain $0 \leq i \leq (a + b)$ where $r(i) = 1$, which means that $i \cdot a = 1 \pmod{a + b}$. This proves also the uniqueness of this factorization.
3. The point Q used to get the palindromic factorization and that refers to the furthest point on the Christoffel path is labeled by $\frac{a+b-1}{b}$. Equivalently, for the point Q , $r(i) = -1$ or we can write: $i \cdot a = -1 \pmod{a + b}$ to determine the exact position of this point.

Example 39 *Let $w = C(\frac{5}{8}) = 0010010100101$ with $n = 13$. From Example 37, we have:*

- $r[8] = 1; w[8] = 1$, then the standard factorization is exactly at position 8 and we get $w[1 \dots 8] = 00100101 = u$ and $w[9 \dots 13] = 00101 = v$ the two words that form the standard factorization of w .
- $r[5] = 12 = n - 1$, then the palindromic factorization is exactly at position 5 and we get $w[1 \dots 5] = 00100 = p_1$ and $w[6 \dots 13] = 10100101 = p_2$ the two palindromes used to obtain the palindromic factorization.

After this definition, we are able to give a supplementary definition for the Christoffel word of slope a/b related to the standard factorization.

Theorem 2.4 [7] *Let w be a binary word over A^* , w is a Christoffel word of slope a/b if and only if we can find two Christoffel words (u, v) such that:*

$$\det \begin{pmatrix} |u|_0 & |v|_0 \\ |u|_1 & |v|_1 \end{pmatrix} \in SL_2(\mathbb{Z}),$$

where $SL_2(\mathbb{Z})$ is the group of invertible matrices in \mathbb{Z} with determinant equal to 1.

The couple (u, v) determines the standard factorization of w .

Example 40 *Let $w = C(\frac{5}{8}) = 0010010100101$, we can write $w = (u, v) = (00100101, 00101)$ where $|u|_0 = 5$; $|u|_1 = 3$ and $|v|_0 = 3$; $|v|_1 = 2$. By computing the determinant as in Theorem 2.4 we get:*

$$\det \begin{pmatrix} |u|_0 & |v|_0 \\ |u|_1 & |v|_1 \end{pmatrix} = \det \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix} = 1.$$

The two Christoffel words $u = C(\frac{3}{5})$ and $v = C(\frac{2}{3})$ form the standard factorization of w .

From ([5], Lemma 3.5), and using Theorem 2.4, we get the Corollary 2.5, where $p \equiv_n q$ represents the congruence between p and q modulo n .

Corollary 2.5 *Let w be a Christoffel word of slope a/b with (u, v) its standard factorization and $n = a + b$. Then $|u| \times a \equiv_n 1$ and $|v| \times b \equiv_n 1$ with $|u| \perp |v|$.*

Example 41 *The standard factorization of $w = C(\frac{5}{8})$ is the pair $(u, v) = (00100101, 00101)$. By Corollary 2.5, the Christoffel words u and v respect the following properties: $8 = |u| \perp |v| = 5$, and $|u| \times 5 = |v| \times 8 = 40 \equiv_{13} 1$.*

Another direct consequence can be obtained from Theorem 2.4. In fact, using the determinants defined in Theorem 2.4, we are able to prove, in a second way, that all the elements of the Christoffel tree are Christoffel words.

Corollary 2.6 *Let w be a Christoffel word with (u, v) its standard factorization. The words $G(w) = (u, uv)$ and $\tilde{D}(w) = (uv, v)$ are Christoffel words.*

Proof. To prove that $G(w)$ and $\tilde{D}(w)$ are Christoffel words, we use Theorem 2.4 and we say:

$$\det \begin{pmatrix} |u|_0 & |uv|_0 \\ |u|_1 & |uv|_1 \end{pmatrix} = \det \begin{pmatrix} |u|_0 & |u|_0 + |v|_0 \\ |u|_1 & |u|_1 + |v|_1 \end{pmatrix} = \det \begin{pmatrix} |u|_0 & |v|_0 \\ |u|_1 & |v|_1 \end{pmatrix} = 1,$$

and since $u, v, w = uv$ are Christoffel words then $G(w)$ is also a Christoffel word. Similarly we prove that $\tilde{D}(w)$ is a Christoffel word too. \square

We know from the previous chapter that we are dealing with circular Christoffel words. As we can see in Definition 2.1, the sequence $(r_i)_{0 \leq i \leq n}$ starts and ends with a 0. Therefore, we are able to represent this Christoffel word on a **Cayley graph** of $\mathbb{Z}/(a+b)\mathbb{Z}$ with generator a . In this graph, the nodes represent the sequence $(r_i)_{0 \leq i \leq |w|}$, where w is the Christoffel word of slope a/b . While the edges are exactly w starting from 0 at angle zero and moving anticlockwise.

Example 42 Let $w = C(\frac{3}{5})$, the sequence $R = [(r_i)_{0 \leq i \leq |w|}]$ is given as follows: $R = [0, 3, 6, 1, 4, 7, 2, 5]$ and $w = 00100101$. Figure 2.3 corresponds to the Cayley graph of this word.

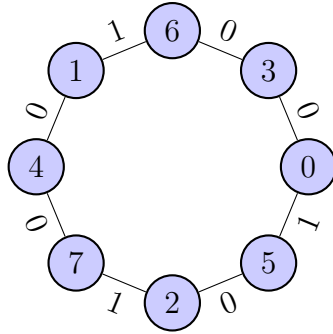


Figure 2.3: The Cayley graph of $w = 00100101$.

We give now a second definition to obtain the Christoffel word w of slope a/b . This definition is based on determining the position of the occurrences of the letter 1 in w . We call the set of these positions the set of **decreasing positions** and we denote it as follows:

Definition 2.7 Let w be a Christoffel word of slope a/b . The set of decreasing positions of w , denoted $D(a, b)$, is: $D(a, b) = \{1 \leq i \leq n \mid w[i] = 1\}$,

The next theorem is a translation of the Corollary 3.2 in [38] given by Paquin and Reutenauer about the set of positions of the Christoffel words. In Chapter 6, we will see in detail the modifications of this corollary in order to get a general form for the set of decreasing positions of a Christoffel word. While for now, we give a simple version of this set and we have:

Theorem 2.8 *Let α be such that $\alpha a \equiv -1 \pmod{n}$, then the set $D(a, b)$ is :*

$$D(a, b) = \{(i\alpha \bmod n) + 1 \mid i = 1 \dots a\}.$$

Note that in the previous equation, one is added to the positions since we consider that words start at position 1 while in [38] the words start at position 0.

Example 43 *Let $a/b = 3/5$ so that $n = 8$, $\alpha = 5$ and $w = 00100101$. The set of the indexes of the letter 1 is $D(3, 5) = \{(5i \bmod 8) + 1 \mid i = 1 \dots 3\} = \{3, 6, 8\}$.*

Axis of symmetry: After defining the Cayley graph and introducing the algebraic definitions of Christoffel words, we try in this paragraph to compute some calculations and determine the position of the axis of symmetry of the Christoffel word on the Cayley graph. By Property 1.13, we know that a Christoffel word can be written as the concatenation of two palindromes and we have $w = p_1 p_2$. From the definition of a palindrome, we can observe that each palindrome has a center of symmetry that is located exactly at the middle of the palindrome. It can be either an element of the word if the length is odd, or it can be located between two values of the word if its length is even. So we let w be a Christoffel word of slope a/b and length n . The words p_1 and p_2 represent the two palindromes obtained from the palindromic factorization of w having ℓ as the position of the palindromic factorization, which means the value that satisfies: $\ell a \equiv_n -1$. Therefore, the positions of the center of symmetries of w are given respectively by: $x_{c_1} = \frac{\ell-1}{2} + 1 = \frac{\ell+1}{2}$ and $x_{c_2} = \frac{\ell+n-1}{2} + 1 = \frac{\ell+n+1}{2}$. Now, we distribute the letters of w and the centers in a proportional way all over the unit circle to retrieve the Cayley graph of w . The position angle on the Cayley graph of each $w[i]$ is $\theta_i = \frac{2\pi \cdot (i-1)}{n}$, $\forall 1 \leq i \leq n$. We arrange the points on the Cayley graph such that the first center of symmetry is always at position $\frac{-\pi}{2}$, hence the position angle of each letter becomes: $\theta'_i = \frac{\pi(4(i-1) - 4x_{c_1} + n)}{2n}$. This order is given for a certain aim that will be developed in Chapter 6. The axis of symmetry of w is the axis joining the two centers of symmetries.

Example 44 *Let $w = C(\frac{7}{5})$ and $n = 12$, we represent w and $R = r[i]$ on the same row and we get:*

$$0 \xrightarrow{0} 7 \xrightarrow{1} 2 \xrightarrow{0} 9 \xrightarrow{1} 4 \xrightarrow{0} 11 \xrightarrow{1} 6 \xrightarrow{1} 1 \xrightarrow{0} 8 \xrightarrow{1} 3 \xrightarrow{0} 10 \xrightarrow{1} 5 \xrightarrow{1} 0$$

Where 0 is for every increasing step and 1 for every decreasing one. By calculating the position of the palindromic factorization, we get $\ell = 5$ where ℓ is obtained by computing: $7\ell \equiv 11 \pmod{12}$. The two centers of symmetry are at positions $x_{c_1} = 3$ and $x_{c_2} = 9$ respectively, and the set of decreasing points is given by: $D(7, 5) = \{2, 4, 6, 7, 9, 11, 12\}$. Figure 2.4 shows the distribution of the letters of the Christoffel word with their position angles on a linear axis. The left part of Figure 2.5 shows the positions of the letters on the Cayley graph starting from angle 0 at position 1, and the right part shows the new positions of the letters after the rotation in order to get the first axis of symmetry at position $-\frac{\pi}{2}$.

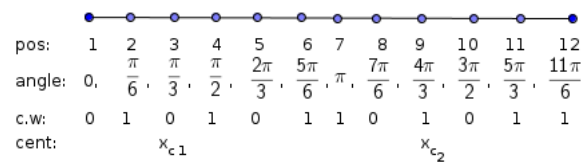


Figure 2.4: The representation of w on a line segment with the position angle of each letter.

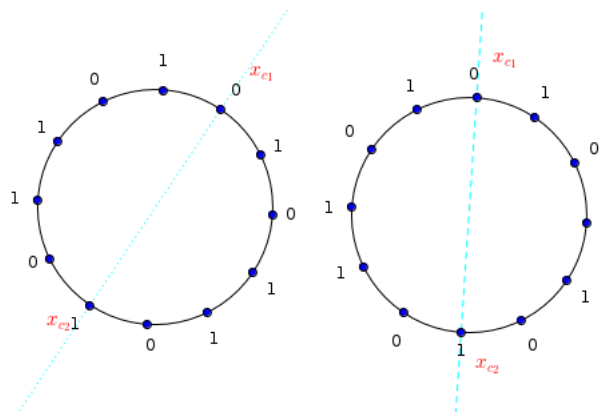


Figure 2.5: The axis of symmetry of $w = C(\frac{7}{5})$ represented on the Cayley graph before and after rotation.

Sage code This is the code used to get the positions of the centers of symmetry for a word w of slope a/b and the position angles of each letter on the Cayley graph.

```

1 import math
2 def pgcd(a,b) :
3     while a%b != 0 :
4         a, b = b, a%b
5     return b
6 def cent(a,n):
7     d=pgcd(a,n)
8     l=1
9     for i in range (n):
10        if (a*i-n+d)%n==0:
11            l=i
12            break
13     xc_1=(l-1)/2.0 +1
14     xc_2=(l-1+n)/2.0 +1
15     ang=[0 for j in range (n)]
16     w=[0 for k in range (n)]
17     ang = [ math.pi * ( 4*(i-1) - 4*xc_1 + n ) / (2*n) for i in range
18 ( n ) ]
19     w = [ 1 if k in s else 0 for k in range(n) ]
20     return (xc_1 ,xc_2 ,ang ,w)

```

Listing 2.1: Coding a word on the Cayley graph with its center of symmetry.

2.2 More about Christoffel words

In the previous Chapter, we defined the trivial Christoffel words to be: $w = 0$ and $w = 1$, where the couple formed by those two elements “(0, 1)” is the root of the Christoffel tree and forms the standard factorization of the Christoffel word of slope 1. Without loss of generality and in order to enlighten the notation, we denote the two endomorphisms G and \tilde{D} by ϕ_0 and ϕ_1 respectively. The vertices of this tree are all the standard factorizations of Christoffel words (see [5], Section 3.2). We recall that using the endomorphism ϕ_0 the next move on the tree is to the left side, while using ϕ_1 , the next move is to the right side. In this section, we introduce some definitions related to the Christoffel words and that are going to be essential tools in the algebraic proves of the main result in this thesis. We start by investigating the directive sequence for the Christoffel words.

Definition 2.9 *Let $w = C(\frac{a}{b})$ be a non-trivial Christoffel word, the directive sequence of w , noted $\Delta(\frac{a}{b})$, is the word $\Delta(\frac{a}{b}) = i_1 i_2 \cdots i_n \in A^*$ such that $w = (\phi_{i_n} \circ \cdots \circ \phi_{i_2} \circ \phi_{i_1})(0, 1)$ and $i_k \in \{0, 1\}$.*

Note that a directive sequence $\Delta(\frac{a}{b}) = i_1 \cdots i_n$ describes the path from the root of the Christoffel tree to the Christoffel word $C(\frac{a}{b})$ as follow : at step k , if

$i_k = 0$ then go left, otherwise, if $i_k = 1$ then go right. More precisely, it allows us to determine the position of the Christoffel word on the Christoffel tree.

Example 45 Let $w = C(\frac{3}{7}) = 0001001001$, we will show using Definition 2.9 that $\Delta(\frac{3}{7}) = 0011$. We start by calculating $(\phi_1 \circ \phi_1 \circ \phi_0 \circ \phi_0)(0, 1)$ and we get:

$$\begin{aligned} (\phi_1 \circ \phi_1 \circ \phi_0 \circ \phi_0)(0, 1) &= (\phi_1 \circ \phi_1 \circ \phi_0)(0, 01) \\ &= (\phi_1 \circ \phi_1)(0, 001) \\ &= \phi_1(0001, 001) \\ &= 0001001001 \\ &= w. \end{aligned}$$

By using $\Delta(\frac{3}{7})$, we are able to place w on the Christoffel tree as shown in Figure 2.6

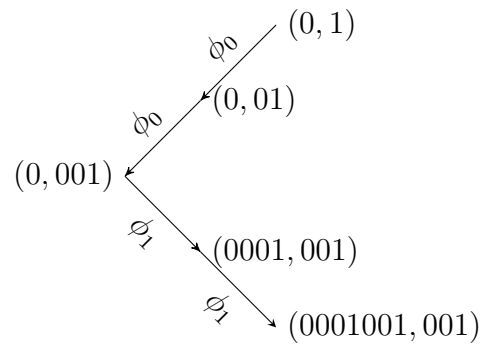


Figure 2.6: Position of $w = C(\frac{3}{7}) = 0001001001$ on the Christoffel tree using $\Delta(\frac{3}{7}) = 0011$.

2.2.1 Continued fractions

The initial geometric definition of Christoffel words mentioned in the last Chapter is “discretization” of a line segment of “rational” slope. Hence, we are not able to study Christoffel words and give results concerning this family of words without giving and understanding some properties of the rational numbers. In this section, we introduce the continued fractions of a rational number a/b . Further more in this chapter, we see the relation between the continued fractions and Christoffel words [4, 7].

Definition 2.10 *The continued fraction of a positive rational number $\frac{a}{b}$ is the finite sequence of integers $\frac{a}{b} = [a_0, \dots, a_z]$, with $a_0 \geq 0$; $a_i \geq 1$ for $1 \leq i \leq z$ and if $z \geq 2$ then $a_z \geq 2$.*

$$\frac{a}{b} = a_0 + \frac{1}{a_1 + \frac{1}{\dots + \frac{1}{a_{z-1} + \frac{1}{a_z}}}}$$

We note that if the sequence is infinite then it is a continued fraction of an irrational number. In this thesis and since we are dealing with Christoffel words that are finite standard sturmian words, we restrict the work to the finite case and we deal with continued fractions of rational numbers.

Example 46 *We consider the rational number $5/7$, and by calculating its continued fraction we get the following form:*

$$\frac{5}{7} = 0 + \frac{1}{1 + \frac{1}{2 + \frac{1}{2}}}, \text{ therefore } \frac{5}{7} = [0, 1, 2, 2].$$

The development in continued fraction can be represented in an equivalent form. The initial definition of the continued fraction imposes to have $a_z \geq 2$ for $z \geq 2$. But for some particular cases, we need to extend the length of the continued fraction or to reduce it. Therefore, Properties 2.11 and 2.12 show us some manipulations in this development keeping the same rational number as result.

Property 2.11 *Let $[a_0, \dots, a_z]$ be the continued fraction of $\frac{a}{b}$, it can also be written as:*

$$\frac{a}{b} = [a_0, a_1, \dots, a_z - 1, 1].$$

Proof. If $\frac{a}{b} = [a_0, a_1, \dots, a_z - 1, 1]$ then

$$\frac{a}{b} = a_0 + \frac{1}{a_1 + \frac{1}{\dots + \frac{1}{a_{z-1} + \frac{1}{1}}}} = a_0 + \frac{1}{a_1 + \frac{1}{\dots + \frac{1}{a_z}}}$$

Hence $\frac{a}{b} = [a_0, a_1, \dots, a_z]$. □

Further in this thesis and during our work, we get $a_i = 0$ for a certain $1 \leq i \leq z$. Since by definition, the a_i 's can not be null for $i \geq 1$, therefore Property 2.12 shows us the modifications to do to get the correct form of the continued fraction.

Property 2.12 Let $\frac{a}{b} = [a_0, a_1, \dots, a_{z-1}, a_z]$, $a_0 \geq 0$ if for $k \in \{1, \dots, z\}$, $a_k = 0$, then $\frac{a}{b} = [a_0, a_1, \dots, a_{k-1} + a_{k+1}, \dots, a_z]$

Proof. Let $\frac{a}{b} = [a_0, a_1, \dots, a_{k-1}, a_k, a_{k+1}, \dots, a_n]$, if $\exists k > 0; a_k = 0$ we get:

$$\frac{a}{b} = a_0 + \frac{1}{a_1 + \frac{1}{\dots + \frac{1}{a_{k-1} + \frac{1}{0 + \frac{1}{a_{k+1} + \frac{1}{\dots + \frac{1}{a_z}}}}}}} = a_0 + \frac{1}{a_1 + \frac{1}{\dots + \frac{1}{a_{k-1} + a_{k+1} + \frac{1}{\dots + \frac{1}{a_z}}}}}.$$

Hence: $\frac{a}{b} = [a_0, a_1, \dots, a_{k-1} + a_{k+1}, \dots, a_z]$ and $a_i > 0, \forall i > 0$. □

In particular if $a_{z-1} = 0$, we get $\frac{a}{b} = [a_0, a_1, \dots, a_{z-2} + a_z]$.

Example 47 The continued fraction $r = [0, 1, 2, 2, 0, 2]$ is written properly as: $r = [0, 1, 2, 4] = \frac{9}{13}$.

$$\frac{9}{13} = 0 + \frac{1}{1 + \frac{1}{2 + \frac{1}{2 + \frac{1}{0 + \frac{1}{2}}}}} = 0 + \frac{1}{1 + \frac{1}{2 + \frac{1}{4}}}$$

The question that arises now is how to obtain the elements of the continued fraction of a rational number? The answer is the recursion used to obtain the a'_i 's of the continued fraction, called partial denominators of the continued fraction of a/b , found in [5, 31] and certainly in Lothaire [33].

Proposition 2.13 Let a/b be a rational number with $[a_0, \dots, a_z]$ its continued fraction. The partial denominators of a/b are calculated in a recursive way as follows: Let $\beta_0 = \frac{a}{b}$ and $a_0 = \lfloor \frac{a}{b} \rfloor$;

- if $i > 0$ and $a_{i-1} \neq \beta_{i-1}$, then:

$$\beta_i = \frac{1}{\beta_{i-1} - a_{i-1}} \quad \text{and} \quad a_i = \lfloor \beta_i \rfloor.$$

- if $i > 0$ and $a_{i-1} = \beta_{i-1}$, then the recursion aborts and we have all the partial denominators of the continued fraction.

Example 48 Let $\frac{a}{b} = \frac{9}{13}$, we obtain the partial denominators of a/b by letting:

$$\begin{aligned} \beta_0 &= \frac{9}{13} & a_0 &= \lfloor \frac{9}{13} \rfloor = 0; \\ \beta_1 &= \frac{1}{\beta_0 - a_0} = \frac{13}{9} & a_1 &= \lfloor \frac{13}{9} \rfloor = 1; \\ \beta_2 &= \frac{1}{\beta_1 - a_1} = \frac{9}{4} & a_2 &= \lfloor \frac{9}{4} \rfloor = 2; \\ \beta_3 &= \frac{1}{\beta_2 - a_2} = 4 & a_3 &= \lfloor 4 \rfloor = 4; \end{aligned}$$

Since $\beta_3 = a_3$, the recursion terminates and we get $\frac{9}{13} = [0, 1, 2, 4]$ as shown in Example 47.

2.2.2 Stern-Brocot tree

The Stern–Brocot tree is a tree that was first introduced by the German mathematician Moritz Abraham Stern and the French clockmaker Achille Brocot in the 19th century [1]. We can find in this tree lot of symmetric mathematical structures with important algebraic and combinatorial properties. It was recently reintroduced by Graham et al. [29]. The representation of this tree is related basically to the regular continued fraction expansion of rational numbers and is presented in Raney [41].

In this part, we introduce the *Stern-Brocot tree* as the infinite tree containing all the positive reduced fractions $\frac{a}{b}$. This tree starts with the fraction $\frac{1}{1}$ at level 1 and gives the two fractions $\frac{1}{2}$ and $\frac{2}{1}$ on the second level using the *Farey sum* defined as follows: $\frac{a}{b} \oplus \frac{c}{d} = \frac{a+c}{b+d}$. For the initial cases, we have: $\frac{1}{1} = \frac{0}{1} \oplus \frac{1}{0}$, $\frac{1}{2} = \frac{1}{1} \oplus \frac{0}{1}$ and $\frac{2}{1} = \frac{1}{1} \oplus \frac{1}{0}$. The two fractions $1/0$ and $0/1$ are not represented on the tree since they are not positive reduced fractions.

To simplify the notations, we denote $\frac{a}{b} \oplus \frac{c}{d}$ by $\frac{a+c}{b+d}$ and $\frac{a}{b} \oplus \dots \oplus \frac{a}{b}$ repeated p times by $(\frac{a}{b})^p$.

On each level of the tree, every fraction is of the form $\frac{a+c}{b+d}$ where $\frac{a}{b}$ (resp. $\frac{c}{d}$) is the nearest ancestor above and to the right (resp. left) of $\frac{a+c}{b+d}$. Figure 2.7 represents the first levels of the Stern-Brocot tree.

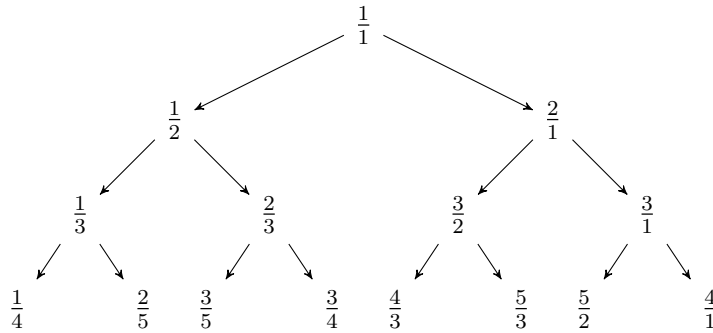


Figure 2.7: The first levels of the Stern-Brocot tree.

Remark 2.14 *By the construction of the tree, we remark the following:*

1. *The tree is divided into two symmetric parts. The left part refers to all the reduced fractions $\frac{a}{b}$ with $a < b$, while the right part is for the reduced fractions of the form $\frac{b}{a}$.*
2. *For each level k , the rational numbers are bounded by $\frac{1}{k}$ and $\frac{k}{1}$.*

3. At a certain level k of the Stern-Brocot tree, the rational numbers appear in an increasing order from the left to the right.
4. At a certain level k , all the rational numbers $\frac{a}{b}$ belonging to the first half of the tree, i.e. ($a < b$), have a denominator greater or equal than k . Consequently, in the second half, the fractions have a numerator greater or equal to k .
5. If $\frac{a}{b}$ and $\frac{c}{d}$ are the two ancestors of $\frac{a+c}{b+d}$, we always have this relation:

$$\frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d}.$$

Example 49 At the fourth level of the Stern-Brocot tree, the following sequence of fractions appear increasingly and respectively:

$$\frac{1}{4} < \frac{2}{5} < \frac{3}{5} < \frac{3}{4} < \frac{4}{3} < \frac{5}{3} < \frac{5}{2} < \frac{4}{1}.$$

If we choose the fraction $\frac{3}{5}$, it has $\frac{1}{2}$ and $\frac{2}{3}$ as ancestors since it can be written as: $\frac{3}{5} = \frac{1}{2} \oplus \frac{2}{3}$. We have by Remark 2.14, $\frac{1}{2} = 0.5 < \frac{3}{5} = 0.6 < \frac{2}{3} = 0.\bar{6}$.

Definition 2.15 For a certain level k of the Stern-Brocot tree, two irreducible fractions a/b and a'/b' are consecutives if they are consecutive after the projection of all the fractions of the tree to level k where the fractions with $b > k$ are excluded. In this case these two fractions respect the following relation:

$$a'b - ab' = 1.$$

We remark that at a certain level k , the consecutive fractions are exactly the ancestors of each fraction at this level.

Property 2.16 The relation between the consecutive fractions, remains true even between $(\frac{a}{b}, \frac{a+c}{b+d})$ and $(\frac{a+c}{b+d}, \frac{c}{d})$.

Proof. Let the couple $(\frac{a}{b}, \frac{a+c}{b+d})$, where $\frac{a}{b} \oplus \frac{c}{d} = \frac{a+c}{b+d}$. The fractions $\frac{a}{b}$ and $\frac{c}{d}$ are the consecutive fractions respecting: $cb - ad = 1$. The same relation holds for this couple since: $(a+c)b - a(b+d) = ab + cb - ab - ad = cb - ad = 1$. Same calculation for the second couple. \square

We can generalize this relation and obtain an interesting relation that will be used during the work in Chapter 7.

Property 2.17 Let $\frac{a}{b}$ and $\frac{c}{d}$ be two consecutive fractions in the Stern-Brocot tree, we have:

$$(c+d)a = c(a+b) - 1.$$

Proof. If a/b and c/d are two consecutive fractions in the Stern-Brocot tree, then:

$$\begin{aligned} cb - ad &= 1 \\ cb - ad - ac + ac &= 1 \\ (a + b)c - (c + d)a &= 1 \\ (a + b)c - 1 &= (c + d)a. \end{aligned}$$

□

Property 2.18 Let $\frac{a}{b}$ and $\frac{c}{d}$ be two consecutive fractions in the Stern-Brocot tree, we have:

$$\frac{c}{d} - \frac{a}{b} = \frac{1}{bd}.$$

Proof. If a/b and c/d are two consecutive fractions in the Stern-Brocot tree, then by Definition 2.15, we have: $cb - ad = 1$. Dividing both sides by bd we get:

$$\begin{aligned} cb - ad &= 1 \\ \frac{cb}{bd} - \frac{ad}{bd} &= \frac{1}{bd} \\ \frac{c}{d} - \frac{a}{b} &= \frac{1}{bd}. \end{aligned}$$

□

Similarly, and since by Property 2.16, we can generalize the result to the couples $(\frac{a}{b}, \frac{a+c}{b+d})$ and $(\frac{a+c}{b+d}, \frac{c}{d})$, then Property 2.17 and Property 2.18 hold also true for these two couples.

Example 50 Let the three fractions: $\frac{2}{7} < \frac{3}{10} < \frac{1}{3}$. From Definition 2.15 and Property 2.16, we have: $7(1) - 2(3) = 7(3) - 2(10) = 1(10) - 3(3) = 1$.

From Property 2.17 and its consequence we have:

- $3(4) = 13(1) - 1 = 12$.
- $2(13) = 9(3) - 1 = 26$.
- $4(2) = 9(1) - 1 = 8$.

From Property 2.18 and its consequence we have:

- $\frac{1}{3} - \frac{3}{10} = \frac{1}{30}$.
- $\frac{1}{3} - \frac{2}{7} = \frac{1}{21}$.
- $\frac{3}{10} - \frac{2}{7} = \frac{1}{70}$.

Due to this tree, a strong link comes out between Combinatorics and arithmetics. In Property 2.19, we give the relation between the Christoffel tree and the Stern-Brocot tree.

Property 2.19 [5] *The Christoffel tree is isomorphic to the Stern-Brocot tree where each vertex of the Christoffel tree of the form (u, v) is associated to the fraction $\frac{|uv|_1}{|uv|_0}$.*

Proof. Let $w = C(\frac{c}{d})$ with (u, v) its standard factorization . We have $\phi_0(u, v) = (u, uv)$ and $w' = \phi_1(u, uv) = (uuv, uv)$. From the construction of the Christoffel tree and its properties, we know that:

- $|u|_1 + |v|_1 = c$ and $|u|_0 + |v|_0 = d$;
- $\phi_0(u, v) = C(\frac{a}{b})$, and we have: $2|u|_1 + |v|_1 = a$ and $2|u|_0 + |v|_0 = b$.

Therefore $w' = \phi_1(u, uv) = (uuv, uv)$ is a Christoffel word having $|w'|_0 = 3|u|_0 + 2|v|_0$ and $|w'|_1 = 3|u|_1 + 2|v|_1$. Which is exactly: $|w'|_1 = a + c$ and $|w'|_0 = b + d$. Hence w' is a Christoffel word of slope $\frac{a+c}{b+d}$. Figure 2.8 shows the position of the Christoffel words isomorphic to their slopes. Same reasoning is repeated if we start by ϕ_1 then by ϕ_0 .

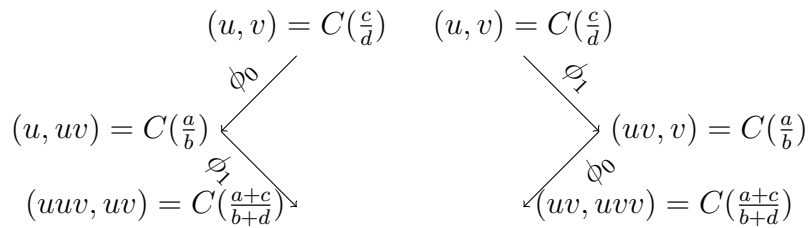


Figure 2.8: Relation between the Farey sum and the endomorphisms ϕ_0 and ϕ_1 .

□

Sage code By considering that the fraction $1/1$ belongs to the first level of the Stern-Brocot tree, we give the code used to get all the fractions of this tree level by level. The fraction a/b is represented by the couple (a, b) .

```

1 def SternBrocot( n ) :
2     allNodes = [ vector((0,1)), vector((1,0)) ]
3     tree = [ allNodes ]
4     for i in range(n) :
5         newNodes = []
6         tmp = []
7         for j in range( len(allNodes)-1 ) :
8             a = allNodes[j]
9             b = allNodes[j+1]
10            newNodes.append( a+b )
11            tmp.append( a )
12            tmp.append( a+b )
13        tmp.append( allNodes[-1] )
14        tree.append( newNodes )
15        allNodes = tmp
16    return tree
17

```

Listing 2.2: Coding used to get the elements of the Stern-Brocot tree represented by couples.

Example 51 *The following couples represent the fractions of the first 3 levels of the Stern-Brocot tree.*

```

1
2 sage: SternBrocot(3)
3
4 [[(0, 1), (1, 0)],
5  [(1, 1)],
6  [(1, 2), (2, 1)],
7  [(1, 3), (2, 3), (3, 2), (3, 1)]]

```

The isomorphism between these two trees creates a link between the continued fractions, the Christoffel words and the directive sequence $\Delta(\frac{a}{b})$. After identifying the partial denominators of the continued fraction of a rational number using Proposition 2.13, we introduce now the following theorem for Henry J. Smith in 1876. This theorem gives an additional characterization for the Christoffel words. In fact, he showed that the Christoffel word can be obtained by a recursive expression using the partial denominators of the rational number.

Theorem 2.20 [49] *Let $\frac{a}{b} = [c_0, c_1 \dots c_n]$, a word $w = 0u1$ is a Christoffel word of slope $\frac{a}{b}$ if and only if $u01$ or $u10$ is equal to s_{n+1} , where s_n is defined recursively by: $s_{-1} = 0, s_0 = 1$ and $s_{n+1} = s_n^{c_n} s_{n-1}$ for all $n \geq 0$.*

Example 52 For the rational number $\frac{9}{13} = [0, 1, 2, 4]$, we can write:

$$\begin{aligned} s_1 &= s_0^0 s_{-1} = 1^0 0 = 0 \\ s_2 &= s_1^1 s_0 = 0^1 1 = 01 \\ s_3 &= s_2^2 s_1 = (01)^2 0 = 01010 \\ s_4 &= s_3^4 s_2 = (01010)^4 01 = 0101001010010100101001. \end{aligned}$$

We remark that s_4 can be written as: $s_4 = (01010010100101001010)01$ where $u = 01010010100101001010$ is a palindrome and $w = 0u1$ is the Christoffel word of slope $\frac{9}{13}$.

Palindromization

In this part, we define an operator on binary words that will be the key chain between rational numbers and the Christoffel word. We start by recalling that a palindrome is the word equal to its reversal and we define the *palindromic closure* as follows:

Definition 2.21 [4] Given a word w , the palindromic closure w^+ is the shortest palindrome having w as a prefix.

Example 53 Let $w = 011010010$, this word is not a palindrome since $\tilde{w} = 010010110 \neq w$. Therefore, we can calculate its shortest palindrome and we get: $w^+ = 011010010\underline{110}$.

Using this operator, we introduce the function *iterated palindromic closure* denoted $Pal(w)$ that is used to construct the Christoffel word.

Definition 2.22 The iterated palindromic closure, is the function $Pal : A^* \rightarrow A^*$ defined recursively as follows:

$$Pal(w) = \begin{cases} \epsilon & \text{if } w = \epsilon, \\ (Pal(v)a)^+ & \text{if } w = va \ \forall a \in A. \end{cases}$$

N.B: Do not mix between $Pal(A^*)$ the set of all palindromes over the alphabet A^* and $Pal(w)$ the iterated palindromic closure function defined on words.

Example 54 Let $w = 0110$, to calculate $Pal(w)$ we must follow these steps:

$$\begin{aligned} Pal(0) &= 0, \\ Pal(01) &= (Pal(0)1)^+ = 010, \\ Pal(011) &= (Pal(01)1)^+ = 01010, \\ Pal(0110) &= (Pal(011)0)^+ = 0101001010. \end{aligned}$$

and we get: $Pal(0110) = \underline{0101001010}$ where each underlined letter belongs to w .

We know that each Christoffel word w is written as: $w = 0w'1$, where w' is the central word and moreover w' is a palindrome. In 1994, Mignosi and Aldo de Luca introduced a property for the Christoffel words that was a path that led to the relation between these words and the directive sequence.

Proposition 2.23 [17] *Any Christoffel word can be written in the following form $w = 0u1$ where $u = \text{Pal}(v)$ for some $v \in \{0, 1\}^*$.*

Example 55 *Let the Christoffel word $w = C(\frac{5}{7})$, we let $v = 0110$. We know that $w = 0(0101001010)1$, hence by Proposition 2.23, $v = \text{Pal}(0110) = 0101001010$, that was calculated in Example 54. In the Figure 2.9, we show the path followed in the Christoffel tree to reach w that corresponds to the same path followed on the Stern-Brocot tree to reach the rational number $5/7$.*

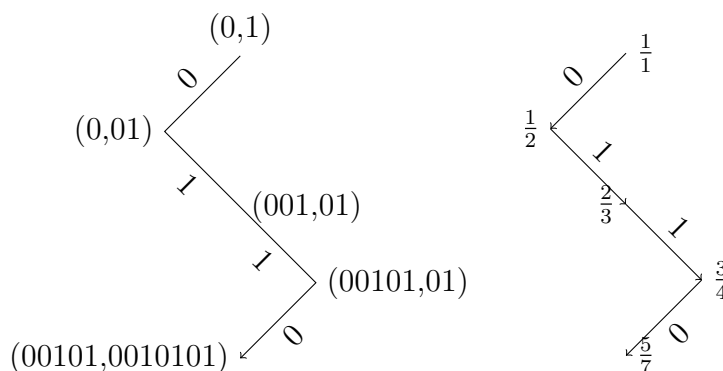


Figure 2.9: Position of the $w = C(\frac{5}{7})$ comparing to the position of $5/7$ on Stern-Brocot tree.

In the following part, we will explain how to pass from the continued fraction of a rational number a/b to the Christoffel word of slope a/b passing by the iterated palindromic closure function. In fact, in 1987, Berstel introduced the following relation between the partial denominators of a rational number and two particular matrices based on the theorem 2.1 in George Raney's paper [41] in 1973. Using the two matrices $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, and for the particular case of the rational number a/b where $a < b$, we have:

Proposition 2.24 [3] *For $a < b$, we associate for the rational number $\frac{a}{b} = [0, a_1, a_2, \dots, a_n]$, the following matrix:*

$$M\left(\frac{a}{b}\right) = A^{a_1-1} B^{a_2} \dots C^{a_n-1},$$

where $C = A$ or $C = B$ depending on the parity of n . We have:

$$M\left(\frac{a}{b}\right) \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} b-a \\ a \end{pmatrix}.$$

Example 56 For the rational number $\frac{9}{13} = [0, 1, 2, 4]$, we have:

1. The associated matrix to $\frac{9}{13}$ is given by: $M\left(\frac{9}{13}\right) = A^0 B^2 A^3 = \begin{pmatrix} 1 & 3 \\ 2 & 7 \end{pmatrix}$.

2. By calculating $M\left(\frac{9}{13}\right) \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, we get: $\begin{pmatrix} 4 \\ 9 \end{pmatrix}$.

Furthermore, in 1993, Borel and Laubie, followed by de Luca in 1997 then Berthé, de Luca and Reutenauer in 2008, gave the following theorem, where they linked these two particular matrices to the Christoffel word of slope a/b .

Theorem 2.25 [4, 6, 7] Let w be a Christoffel word of slope a/b , there exists a unique palindrome v such that $w = 0\text{pal}(v)1$, if (w_1, w_2) is the standard factorization of w then:

$$\mu(v) = \begin{pmatrix} |w_1|_0 & |w_2|_0 \\ |w_1|_1 & |w_2|_1 \end{pmatrix},$$

where the multiplicative monoid morphism

$$\mu : \{0, 1\}^* \longrightarrow SL_2(\mathbb{Z})$$

is defined by : $\mu(0) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = A$ and $\mu(1) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = B$.

Example 57 For the rational number $\frac{3}{11} = [0, 3, 1, 2]$, with $w = C\left(\frac{3}{11}\right) = 00001000010001$ the Christoffel word of slope $3/11$ and $w = (00001, 000010001)$ its standard factorization, we let $v = 00010$ and we have:

$$\mu(v) = \mu(0^3 1 0) = \mu(0)^3 \mu(1) \mu(0) = A^3 B^1 A^1 = \begin{pmatrix} 4 & 7 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} |w_1|_0 & |w_2|_0 \\ |w_1|_1 & |w_2|_1 \end{pmatrix}.$$

Using Proposition 2.24 and Theorem 2.25, we can conclude the general form of the directive sequence of a rational number a/b and the link between $\Delta\left(\frac{a}{b}\right)$ and the Christoffel word of slope a/b .

Theorem 2.26 Let w be a Christoffel word of slope $\frac{a}{b}$ where the continued fraction of a/b is given by: $[a_0, a_1, \dots, a_z] = [a_0, a_1, \dots, a_z - 1, 1]$. The directive sequence of this Christoffel word has the following form:

$$\Delta\left(\frac{a}{b}\right) = 1^{a_0} 0^{a_1} 1^{a_2} \dots p^{a_z - 1}$$

where $p \in \{0, 1\}$. In addition, we have: $w = 0\text{Pal}(\Delta\left(\frac{a}{b}\right))1$.

Proof. We divide the proof into two parts, the first one where $a_0 = 0$ and the second one for $a_0 > 0$.

- For $a_0 = 0$:

From Theorem 2.25, the Christoffel word of slope a/b has a unique word v such that: $w = 0Pal(v)1$ with $\mu(v) = \begin{pmatrix} |w_1|_0 & |w_2|_0 \\ |w_1|_1 & |w_2|_1 \end{pmatrix}$. In addition, from Proposition 2.24, for the rational number $\frac{a}{b} = [0, a_1, \dots, a_z]$, we have: $M(\frac{a}{b}) = A^{a_1-1}B^{a_2} \dots C^{a_z-1}$ where $C = A$ or $C = B$ depending on the parity of z and with $M(\frac{a}{b}) \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} b-a \\ a \end{pmatrix}$. If we multiply by A from the left side of the matrix $M(\frac{a}{b})$, we get: $A.M(\frac{a}{b}) = A^{a_1}B^{a_2} \dots C^{a_z-1} = \mu(0^{a_1}1^{a_2} \dots p^{a_z-1})$, where $p = 0$ or $p = 1$ depending on the parity of z .

$$\begin{aligned} A.M(\frac{a}{b}) \begin{pmatrix} 1 \\ 1 \end{pmatrix} &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} b-a \\ a \end{pmatrix} \\ &= \begin{pmatrix} b \\ a \end{pmatrix} \\ &= \begin{pmatrix} |w|_0 \\ |w|_1 \end{pmatrix}. \end{aligned}$$

In another hand, $\mu(v) \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} |w_1|_0 + |w_2|_0 \\ |w_1|_1 + |w_2|_1 \end{pmatrix} = \begin{pmatrix} |w|_0 \\ |w|_1 \end{pmatrix}$. This allows us to conclude that $v = 1^0 0^{a_1} \dots p^{a_z-1}$ where $p \in \{0, 1\}$ depending on the parity of z . By Definition 2.9, we know that the directive sequence of a rational number a/b is the path followed in the Stern-Brocot tree to reach the rational number. Hence we have that $w = 0Pal(\Delta(\frac{a}{b}))1$.

- For $a_0 > 0$,

From Theorem 2.20, we are able to pass from the partial denominators of a rational number $a/b = [a_0, \dots, a_n]$ to the Christoffel word of slope a/b by using the smith algorithm. In fact the word obtained is of the form $s_{n+1} = u01$ or $s_{n+1} = u10$ where u is a palindrome and $s_{n+1} = s_n^{a_n} s_{n-1}$. By Theorem 2.25, we know that the Christoffel word has the following form: $0Pal(v)1$. Therefore, we can conclude that $u = Pal(v)$. Starting from $s_1 = s_0^{a_0} s_{-1} = 1^{a_0} 0$, if we compute $s_2 = s_1^{a_1} s_0 = (1^{a_0} 0)^{a_1} 1$ and we continue the iteration till s_{n+1} , we remark that u always starts with 1^{a_0} . Hence, using the iterated palindromic closure function and since $Pal(1^{a_0}) = 1^{a_0}$, we can say that v starts with 1^{a_0} . Finally, from the previous case, we can generalize the form of $\Delta(\frac{a}{b})$ and say that $v = \Delta(\frac{a}{b}) = 1^{a_0} 0^{a_1} \dots p^{a_n-1}$, where $p \in \{0, 1\}$ depending on the parity of n and $w = 0Pal(v)1 = 0Pal(\Delta(\frac{a}{b}))1$.

□

Example 58 Let us consider the rational number $5/7$ of continued fraction $5/7 = [0, 1, 2, 2]$.

Using Theorem 2.26, the directive sequence of this fraction is given by: $\Delta(\frac{5}{7}) = 1^0 0^1 1^2 0^1 = 0110$. Hence, we can write: $C(\frac{5}{7}) = 0 \text{Pal}(\Delta(\frac{5}{7}))1 = 0 \boxed{0101001010} 1$ since $\text{Pal}(0110) = 0101001010$ as seen in Example 55.

In the following, we give some properties about the directive sequence of the rational number deduced from Theorem 2.26.

Property 2.27 If the continued fraction of a/b is of the form: $\frac{a}{b} = [0, m, m, \dots, m]$ where m is repeated ℓ times, then $\Delta(\frac{a}{b})$ is of the following form:

$$\Delta\left(\frac{a}{b}\right) = \begin{cases} (0^m 1^m)^{f-1} 0^m 1^{m-1} & \text{if } \ell = 2f \\ (0^m 1^m)^f 0^{m-1} & \text{if } \ell = 2f + 1 \end{cases}$$

Proof. Since $a_0 = 0$ and $1^0 = \varepsilon$ then Δ starts with 0^m . We have $[0, m, m, \dots, m] = [0, m, m, \dots, m-1, 1]$ where m is repeated $\ell - 1$ times and we have $a_\ell = m - 1$. By applying Theorem 2.26 we get:

- If $\ell = 2f + 1$, then we have $2f$ times the number m and the $m - 1$ only one time therefore $\Delta = (0^m 1^m)^f 0^{m-1}$.
- If $\ell = 2f$, then we have the m repeated $2(f - 1)$ times followed by 1 (m) and 1 ($m - 1$) therefore $\Delta = (0^m 1^m)^{f-1} 0^m 1^{m-1}$.

□

In the next Chapter, we need some particular fractions on the Stern-Brocot tree. Therefore, we define the following terminology by reference to the position of the fractions, relatively to a/b , in the Stern-Brocot tree. See Figure 2.10 for an illustration. We give the definition of those particular fractions in two different ways. First by introducing their form using the continued fraction then deducing their directive sequence. Second, we give the definition in the opposite way by introducing the form of the directive sequence and deducing the continued fraction form of each fraction showing geometrically the paths needed to get each fraction.

Definition 2.28 Given $\frac{a}{b} = [a_0, \dots, a_z]$,

- (1) the top branch fraction of $\frac{a}{b}$, noted $TBF(\frac{a}{b})$, is the fraction $[a_0, \dots, a_{z-1} + 1]$,
- (2) the first reduced fraction of $\frac{a}{b}$, noted $FRF(\frac{a}{b})$, is the fraction $[a_0, \dots, a_{z-1}]$,
- (3) the first extended fraction of $\frac{a}{b}$, noted $FEF(\frac{a}{b})$, is the fraction $[a_0, a_1, \dots, a_z + 1]$,
- (4) the first deviation fraction of $\frac{a}{b}$, noted $FDF(\frac{a}{b})$, is the fraction $[a_0, a_1, \dots, a_z - 1, 2]$,
- (5) the first parallel fraction of $\frac{a}{b}$, noted $FPF(\frac{a}{b})$, is either $[a_0, a_1, \dots, a_{z-1} - 1, 2]$ if $a_{z-1} \neq 1$ or $[a_0, a_1, \dots, a_{z-2} + 2]$ if $a_{z-1} = 1$.
- (6) the second unidirectional father of $\frac{a}{b}$, noted $SUF(\frac{a}{b})$, is either $[a_0, a_1, \dots, a_z - 2]$ if $a_z > 2$ or $[a_0, a_1, \dots, a_{z-2}]$ if $a_z = 2$.

Note that $SUF(\frac{a}{b})$ is not defined for fractions $\frac{1}{1}$, $\frac{1}{2}$ and $\frac{2}{1}$. Using Theorem 2.26 we have the following relations between the directive sequence of a fraction and the above definitions.

Remark 2.29 Given $\frac{a}{b} = [a_0, a_1, \dots, a_z]$, we have that $\Delta(\frac{a}{b}) = 1^{a_0}0^{a_1} \dots p^{a_{z-1}}q^{a_z-1}$ where $\{p, q\} = \{0, 1\}$ and:

- (1) $\Delta(TBF(\frac{a}{b})) = 1^{a_0}0^{a_1} \dots p^{a_{z-1}}$,
- (2) $\Delta(FRF(\frac{a}{b})) = 1^{a_0}0^{a_1} \dots p^{a_{z-1}-1}$,
- (3) $\Delta(FEF(\frac{a}{b})) = 1^{a_0}0^{a_1} \dots p^{a_{z-1}}q^{a_z}$,
- (4) $\Delta(FDF(\frac{a}{b})) = 1^{a_0}0^{a_1} \dots p^{a_{z-1}}q^{a_z-1}p$,
- (5) $\Delta(FPF(\frac{a}{b})) = 1^{a_0}0^{a_1} \dots p^{a_{z-1}-1}q$ if $a_{z-1} \neq 1$, or $1^{a_0}0^{a_1} \dots q^{a_z-2+1}$ if $a_{z-1} = 1$.
- (6) $\Delta(SUF(\frac{a}{b})) = 1^{a_0}0^{a_1} \dots q^{a_z-1}$ if $a_z > 2$, or $1^{a_0}0^{a_1} \dots q^{a_z-2-1}$ if $a_z = 2$.

Example 59 Let $a/b = 3/5$, its continued fraction is : $[0, 1, 1, 2]$, applying the previous result, we get: $TBF(3/5) = 2/3$, $FEF(3/5) = 4/7$, $FDF(3/5) = 5/8$, $FPF(3/5) = 1/3$ and $SUF(3/5) = 1/1$. See Figure 2.10 for the positions of these fractions in the Stern-Brocot tree.

Definition 2.30 Let a/b be a rational number with $\Delta(\frac{a}{b}) = up$ where $u \in A^*$ and $p \in \{0, 1\}$, we define the first extended fraction of a/b as the fraction c/d such that $\Delta(\frac{c}{d}) = upp$.

Property 2.31 Let $\frac{a}{b} = [a_0, a_1, \dots, a_z]$ then the first extended fraction of $\frac{a}{b}$ is $\frac{c}{d} = [a_0, a_1, \dots, a_z + 1]$.

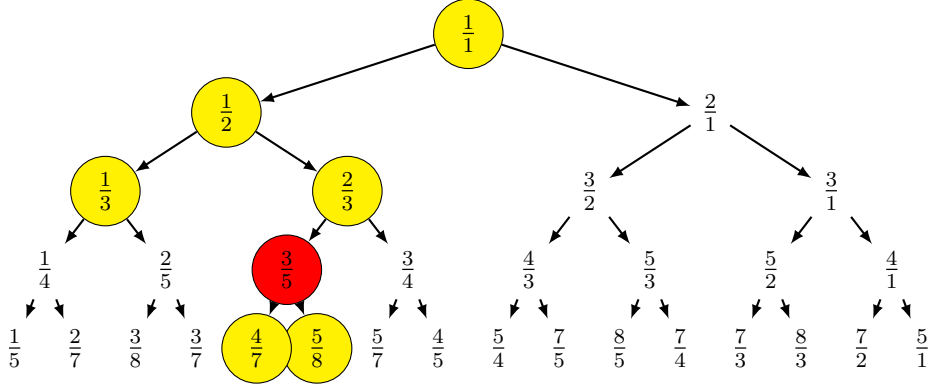
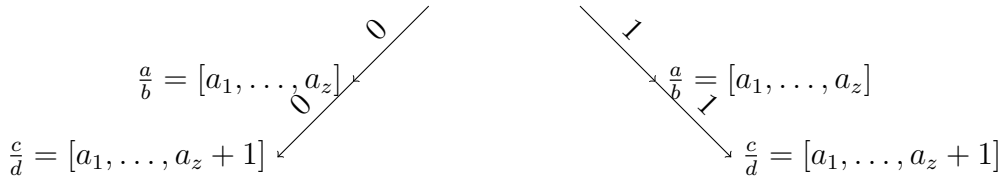


Figure 2.10: Illustration of Definition 2.28 in the Stern-Brocot tree relatively to $3/5 = [0, 1, 1, 2]$. We have : $\text{TBF}(3/5) = 2/3$, $\text{FRF}(3/5) = 1/2$, $\text{FEF}(3/5) = 4/7$, $\text{FDF}(3/5) = 5/8$, $\text{FPF}(3/5) = 1/3$ and $\text{SUF}(3/5) = 1/1$.

Proof. Let $\frac{a}{b} = [a_0, \dots, a_z] = [a_0, \dots, a_z - 1, 1]$ with $\Delta(\frac{a}{b}) = 1^{a_0} 0^{a_1} \dots p^{a_z-1} = u'p^{a_z-2}p$ where $u = u'p^{a_z-2}$ and $p = 0$, (resp. $p = 1$) if ϕ_0 , (resp. ϕ_1) is the last morphism used to reach a/b on the Stern-Brocot tree. By definition 2.30, $\Delta(\frac{c}{d}) = u'p^{a_z-2}pp = u'p^{a_z}$. Thus $\frac{c}{d} = [a_0, \dots, a_z, 1] = [a_0, \dots, a_z + 1]$. \square



Definition 2.32 Let a/b be a rational number with $\Delta(\frac{a}{b}) = up$ where $u \in A^*$ and $p \in \{0, 1\}$, we define the first deviation fraction of a/b as the fraction c/d such that $\Delta(\frac{c}{d}) = upp'$, where $p' \in \{0, 1\}$ and $p \neq p'$.

Lemma 2.33 Let $\frac{a}{b} = [a_0, a_1, \dots, a_z]$, then the first deviation fraction of $\frac{a}{b}$ is $\frac{c}{d} = [a_0, \dots, a_z - 1, 2]$.

Proof. Let $\frac{a}{b} = [a_0, a_1, \dots, a_z]$ be a rational number on the Stern-Brocot tree, then $\Delta(\frac{a}{b}) = 1^{a_0} \dots p^{a_z-1} = u'p^{a_z-2}p$ where $u' \in A^*$, $u = u'p^{a_z-2}$ and $p = 0$, resp $p = 1$ if ϕ_0 , resp ϕ_1 is the last morphism used to reach a/b . By definition 2.32, $\Delta(\frac{c}{d}) = 1^{a_0} \dots p^{a_z-1}p'$ where $p' \in \{0, 1\}$ and $p' \neq p$. Thus $\frac{c}{d} = [a_0, \dots, a_z - 1, 1, 1] = [a_0, \dots, a_z - 1, 2]$.

$$\begin{array}{ccc}
\frac{a}{b} = [a_1, \dots, a_z] & \begin{array}{c} \swarrow 0 \\ \searrow 1 \end{array} & \\
\frac{c}{d} = [a_1, \dots, a_z - 1, 2] & & \\
\end{array}
\qquad
\begin{array}{ccc}
\begin{array}{c} \swarrow 1 \\ \searrow 0 \end{array} & \frac{a}{b} = [a_1, \dots, a_z] & \\
\frac{c}{d} = [a_1, \dots, a_z - 1, 2] & & \\
\end{array}$$

□

Definition 2.34 Let a/b be a rational number with $\Delta(\frac{a}{b}) = upp'$ where $u \in A^*$, $p, p' \in \{0, 1\}$ and $p \neq p'$. We define the first parallel fraction of a/b as the fraction c/d such that $\Delta(\frac{c}{d}) = up'$.

Lemma 2.35 Let $\frac{a}{b} = [a_0, a_1, \dots, 2]$, then the first parallel fraction of a/b is c/d and is equal to:

$$\frac{c}{d} = \begin{cases} [a_0, \dots, a_{z-1} - 1, 2] & \text{if } a_{z-1} \neq 1 \\ [a_0, \dots, a_{z-2} + 2] & \text{if } a_{z-1} = 1 \end{cases} \quad (2.1)$$

Proof. Let $\frac{a}{b} = [a_0, \dots, a_{z-1}, 2]$ be a rational number with $\Delta(\frac{a}{b}) = u'p^{a_{z-1}-1}pp'$ where $u' \in A^*$, $p, p' \in \{0, 1\}$ and $p \neq p'$. By definition 2.34, the first parallel fraction c/d can be written as: $\Delta(\frac{c}{d}) = u'p^{a_{z-1}-1}p'$. Therefore, if $a_{z-1} \neq 1$ we get $\frac{c}{d} = [a_0, \dots, a_{z-1} - 1, 2]$. If $a_{z-1} = 1$ and since we can write $u' = u''p^{a_{z-2}}$, where $u'' \in A^*$, we get $\Delta(\frac{c}{d}) = u''p^{a_{z-2}}p' = u''p^{a_{z-2}+1}$. Thus $\frac{c}{d} = [a_0, \dots, a_{z-2} + 2]$. □

$$\begin{array}{ccc}
[a_0, \dots, a_{z-1}] = [a_1, \dots, a_{z-1} - 1, 1] & \begin{array}{c} \swarrow 1 \\ \searrow 0 \end{array} & \\
\frac{c}{d} = [a_0, \dots, a_{z-1} - 1, 2] & & [a_0, \dots, a_{z-1}, 1] = [a_1, \dots, a_{z-1} + 1] \\
\frac{a}{b} = [a_1, \dots, a_{z-1}, 2] & & \\
\end{array}$$

or

$$\begin{array}{ccc}
[a_0, \dots, a_{z-1}] = [a_1, \dots, a_{z-1} - 1, 1] & \begin{array}{c} \swarrow 1 \\ \searrow 0 \end{array} & \\
[a_0, \dots, a_{z-1}, 1] = [a_1, \dots, a_{z-1} + 1] & & \frac{c}{d} = [a_0, \dots, a_{z-1} - 1, 2] \\
\frac{a}{b} = [a_1, \dots, a_{z-1}, 2] & & \\
\end{array}$$

Definition 2.36 Let a/b be a rational number with $\Delta(\frac{a}{b}) = uv$; $u \in A^*$ and v is the shortest suffix of $\Delta(\frac{a}{b})$; $|v|_p = k$ and $p \in \{0, 1\}$. We define the fraction c/d as the fraction k steps before a/b and in the same direction of it, if $\Delta(\frac{c}{d}) = u$.

Lemma 2.37 Let $\frac{a}{b} = [a_0, a_1, \dots, a_z]$, then the fraction c/d which is k steps before a/b is equal to:

$$\frac{c}{d} = \begin{cases} [a_0, \dots, a_z - k] & \text{if } a_z > k \\ [a_0, \dots, a_{z-2}] & \text{if } a_z = k. \end{cases} \quad (2.2)$$

Proof. Let $\frac{a}{b} = [a_0, \dots, a_{z-2}, a_{z-1}, a_z]$ with $\Delta(\frac{a}{b}) = u'p^{a_z-2}p'^{a_z-1}p^{a_z-1}$ where $p \neq p'$ and $p, p' \in \{0, 1\}$. By definition 2.36, if $a_z > k$, we choose $v = p^k$ and we write $\Delta(\frac{a}{b}) = uv$, hence $\Delta(\frac{c}{d}) = u = u'p^{a_z-2}p'^{a_z-1}p^{a_z-1-k}$ and $\frac{c}{d} = [a_0, \dots, a_z - k]$. If $a_z = k$ we choose $v = pp'^{a_z-1}p^{a_z-1}$ hence $\Delta(\frac{c}{d}) = u = u'p^{a_z-2-1}$ thus $\frac{c}{d} = [a_0, \dots, a_{z-2}]$. \square

In the last part of this Chapter and using the definition of Paquin, Reutenauer for Christoffel words that we saw in Definition 2.8 [38], we give some additional arithmetic properties for the Balanced matrix and an equivalent arithmetic construction for $B_{\frac{a}{b}}$.

We recall that: $B_{\frac{a}{b}}[i, j] = S_{\frac{a}{b}}[i, j] - \min_i(S_w[i])$, for all $1 \leq i \leq (a + b)$.

Lemma 2.38 Let $\frac{a}{b} = [a_0, \dots, a_z]$, and $\text{FRF}(\frac{a}{b}) = \frac{x}{y}$ with $|\frac{x}{y}| = k$. We recall the set of decreasing positions $D(a, b) = \{i\alpha + 1 \pmod n\}$ where $n = a + b$ and we have:

$$\alpha = \begin{cases} k & \text{if } z \text{ is even} \\ n - k & \text{if } z \text{ is odd.} \end{cases}$$

Proof. Let $\frac{a}{b} = [a_0, \dots, a_z]$ and $D(a, b) = \{i\alpha + 1 \pmod n\}$ where $\alpha.a \equiv_n -1$. Using Lemma 2.5, the $C(\frac{a}{b}) = (w_1, w_2)$ is such that $|w_1|.a \equiv_n 1 \Leftrightarrow (n - |w_1|).a \equiv_n -1$. Therefore in $D(a, b)$, we have $\alpha = n - |w_1| = |w_2|$.

- If z is even, then $\Delta(\frac{a}{b}) = 1^{a_0}0^{a_1} \dots 1^{a_z-1}$ and the last function used is ϕ_1 which means that the last path used to reach the fraction $\frac{a}{b}$ on the Stern-Brocot tree is the right path. By the definitions of the two functions ϕ_0 and ϕ_1 , $C(\frac{a}{b})$ is obtained by a concatenation to the left side of the word as shown in Figure 2.11.

where $w_2 = C(\text{FRF}(\frac{a}{b}))$ and $|w_2| = k$, therefore $\alpha = k$.

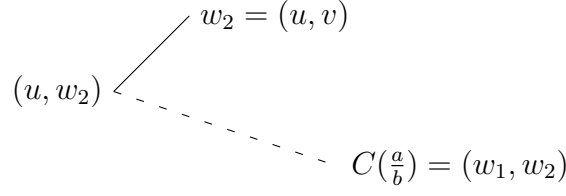


Figure 2.11: Form of the Christoffel word $C(\frac{a}{b})$ obtained by ϕ_1 .

- If z is odd, then $\Delta(\frac{a}{b}) = 1^{a_0}0^{a_1} \dots 0^{a_z-1}$ and the last function used is ϕ_0 which means that the last path used to reach the fraction $\frac{a}{b}$ on the Stern-Brocot tree is the left path. In this case the concatenation is on the right side as shown in Figure 2.12.

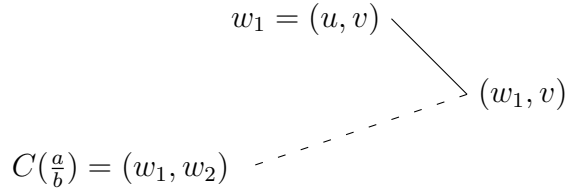


Figure 2.12: Form of the Christoffel word $C(\frac{a}{b})$ obtained by ϕ_0 .

where $w_1 = C(\text{FRF}(\frac{a}{b}))$ and $|w_1| = k$, therefore $\alpha = n - k$.

□

Lemma 2.39 Let $\frac{a}{b} = [a_0, \dots, a_z]$ and $\frac{x}{y} = \text{FRF}(\frac{a}{b})$ such that $|C(\frac{x}{y})| = x + y = k$ and if z is odd, we have:

$$\begin{cases} B_{\frac{a}{b}}[k] = 0^\alpha 10^{n-\alpha-1} \\ B_{\frac{a}{b}}[2k] = 0^{n-2k} 10^{k-1} 10^{k-1}. \end{cases} \quad (2.3)$$

while if z is even, we have:

$$\begin{cases} B_{\frac{a}{b}}[n-k] = 0^\alpha 10^{n-\alpha-1} \\ B_{\frac{a}{b}}[n-2k] = 0^{n-2k} 10^{k-1} 10^{k-1}. \end{cases} \quad (2.4)$$

Proof. We prove the case where z is odd and the second case is obtained in a similar way. We let p_1, p_2 be two palindromes such that: $C(\frac{s}{t}) = 0p_11, C(\frac{g}{h}) = 0p_21$ where $C(\frac{x}{y}) = C(\frac{g}{h})C(\frac{s}{t})$; by composing $(a_z - 1)$ the endomorphism ϕ_1 , we can write: $C(\frac{a}{b}) = (C(\frac{x}{y}))^{a_z-1}C(\frac{u}{v})$ and $C(\frac{u}{v}) = C(\frac{x}{y})C(\frac{s}{t})$. The central word of $C(\frac{x}{y})$ is a palindrome then we have: $p_210p_1 = p_101p_2$. By the definition of the balance matrix, we have for all $1 \leq i \leq n - 1$ and $1 \leq j \leq n$:

$$\begin{aligned} B_{\frac{a}{b}}[k, j] &= S_{\frac{a}{b}}[k, j] - \min(S_{\frac{a}{b}}[k]) \\ &= |C\left(\frac{a}{b}\right)[j \dots k + j - 1]|_1 - \min(S_{\frac{a}{b}}[k]). \end{aligned}$$

Summing up all this we get:

$$\begin{aligned} C\left(\frac{a}{b}\right) &= C\left(\frac{x}{y}\right)^{a_z-1} \cdot C\left(\frac{x}{y}\right) C\left(\frac{s}{t}\right) \\ &= C\left(\frac{x}{y}\right)^{a_z-1} \cdot 0P_210p_110p_11 \\ &= C\left(\frac{x}{y}\right)^{a_z-1} \cdot 0P_101p_210p_11. \end{aligned}$$

We know that $k = |p_1| + |p_2| + 4$ and $|C\left(\frac{x}{y}\right)|_1 = |p_1|_1 + |p_2|_1 + 2 = x$. Therefore all the elements $u \in L_k(C\left(\frac{a}{b}\right)^\omega)$ have $|u|_1 = |p_1|_1 + |p_2|_1 + 2$ except $u' = 1p_210p_11$ who is of length k and at position $n - k$ but has $|u'|_1 = x + 1$. Hence by subtracting by $\min(S_{\frac{a}{b}}[k])$, we get $B_{\frac{a}{b}}[k] = 0^{n-k}10^{k-1}$. By lemma 2.38, we know that since z is odd, $\alpha = n - k$, which ends the proof.

Same reasoning is used to determine the row $B_{\frac{a}{b}}[2k]$, where $2k = 2|C\left(\frac{x}{y}\right)| = 2|p_1| + 2|p_2| + 8$. We also have:

$$\begin{aligned} C\left(\frac{a}{b}\right) &= C\left(\frac{x}{y}\right)^{a_z-2} \cdot C\left(\frac{x}{y}\right) \cdot C\left(\frac{x}{y}\right) C\left(\frac{s}{t}\right) \\ &= C\left(\frac{x}{y}\right)^{a_z-2} \cdot 0P_210p_11 \cdot 0P_210p_110p_11 \\ &= C\left(\frac{x}{y}\right)^{a_z-2} \cdot 0P_101p_21 \cdot 0P_101p_210p_11. \end{aligned}$$

All the elements u of $L_{2k}(C\left(\frac{a}{b}\right)^\omega)$ have $|u|_1 = 2|p_1|_1 + 2|p_2|_1 + 4 = 2x$ except the element $u' = 1p_21 \cdot 0P_101p_210p_11$ of length $2k$ and that appears two times in $C\left(\frac{a}{b}\right)^\omega$, at position $n - k$ and $n - 2k$. We have $|u'|_1 = 2|p_1|_1 + 2|p_2|_1 + 5 = 2x + 1$, thus applying the definition of $B_{\frac{a}{b}}[2k, j] \forall 1 \leq j \leq n$, we obtain the result. \square

2.2.3 Second construction of the balance test matrix

Let w be a Christoffel word, we build the matrix B_w by computing recursively all the positions of its 1's. Let $D_i = \{j \mid B_w[i, j] = 1\}$ that is the set of the positions of the 1's on the i^{th} row. Note that these positions are considered in the circular word of length n so that $j + n = j$ and $1 \leq j \leq n$. For any set of integers S and any integer k , we denote $S + k = \{a + s \mid a \in S\}$.

Theorem 2.40 *If $a < b$ then the sets $(D_i)_{1 \leq i \leq n-1}$ are obtained recursively by: $D_1 = D(a, b)$ and for each i from 2 to $n - 1$:*

$$D_i = \begin{cases} D_{i-1} \cup (D_1 - (i - 1)) & \text{if } i \notin D_1 \\ D_{i-1} \cap (D_1 - (i - 1)) & \text{if } i \in D_1 \end{cases} \quad (2.5)$$

Proof. From the definition of $D(a, b)$ and since $B_w[1] = w$, we have $D_1 = D(a, b)$. By construction, $|D_1| = a$ and we write $D_1 = \{d_1, d_2, \dots, d_a\}$, where $d_1 < d_2 < \dots < d_a$.

Since $a < b$ the Christoffel word is the concatenation of blocks of the form: $0^{d+1}1$ and 0^d1 where the first block is always of the form $0^{d+1}1$. We define an *intersection row* D_i to be a row such that $i \in D_1$. First, we show that the intersection rows are exactly the rows where the values of $\min(S_w[i])$ and $\max(S_w[i])$ increase.

Let D_i be an intersection row, since there is one more occurrence of the letter 1 in the prefix of length i than in the prefix of length $i - 1$, we have $S_w[i, 1] = S_w[i - 1, 1] + 1$. Lemma 1.15 implies that $\min(S_w[i]) = \min(S_w[i - 1]) + 1$. Since a Christoffel word has no period smaller than its length and is 1-balanced, we must have that $\max(S_w[i]) = \min(S_w[i]) + 1$ for all $i \in \{1, \dots, n - 1\}$. Thus $\max(S_w[i]) = \max(S_w[i - 1]) + 1$ if and only if D_i is an intersection row and $\max(S_w[i]) = \max(S_w[i - 1])$ otherwise. There are two cases to consider.

- If $i \notin D_1$, in this case $\max S_w[i] = \max S_w[i - 1]$ and $\min S_w[i] = \min S_w[i - 1] = S_w[i, 1] = S_w[i - 1, 1]$. Let $j \in D_i$, then by definition:

$$B_w[i, j] = S_w[i, j] - \min S_w[i] = S_w[i, j] - S_w[i, 1] = 1.$$

and

$$B_w[i - 1, j] = S_w[i - 1, j] - \min S_w[i - 1] = S_w[i - 1, j] - S_w[i - 1, 1] \in \{0, 1\}.$$

Indeed, $j \in D_i$ implies $B_w[i, j] = 1$ but $B_w[i - 1, j]$ can be equal to 0 or 1.

- If $B_w[i - 1, j] = 1$, then we have $j \in D_{(i-1)}$.

– If $B_w[i-1, j] = 0$,

$$\begin{aligned} S_w[i-1, j] - S_w[i-1, 1] &= 0, \\ S_w[i-1, j] - S_w[i, 1] &= 0, \\ S_w[i-1, j] - S_w[i, j] + 1 &= 0, \\ S_w[i, j] - S_w[i-1, j] &= 1, \\ |w[j, i+j-1]|_1 - |w[j, i+j-2]|_1 &= 1. \end{aligned}$$

Therefore, $w[i+j-1] = 1$ and $B_w[1, i+j-1] = S_w[1, i+j-1] = w[i+j-1] = 1$, hence $i+j-1 \in D_1 \Leftrightarrow j \in D_1 - (i-1)$.

From the above, we conclude that $j \in D_i \Leftrightarrow j \in D_{(i-1)} \cup (D_1 - (i-1))$.

- If D_i is an intersection row, then $i \in D_1$ and $S_w[i, 1] = S_w[i-1, 1] + 1$;

$$\begin{aligned} j \in D_i \Leftrightarrow S_w[i, j] &= S_w[i, 1] + 1, \\ &= S_w[i-1, 1] + 2. \end{aligned}$$

Let $j \in D_i$, by construction $S_w[i-1, j]$ can be equal to $S_w[i-1, 1]$ or to $S_w[i-1, 1] + 1$. By contradiction, suppose $S_w[i-1, j] = S_w[i-1, 1]$. In this case, $S_w[i, j] = S_w[i-1, j] + 2$, which is impossible since w is 1-balanced. Hence, $S_w[i-1, j] = S_w[i-1, 1] + 1$ and in this case $B_w[i-1, j] = 1$ and $j \in D_{(i-1)}$.

It remains to show that $j \in D_1 - (i-1)$. Since $j \in D_i$ and $B_w[i, j] = 1$, we get $\min(S_w[i]) = (S_w[i, j]) - 1$. Knowing that $\min(S_w[i]) = \min(S_w[i-1]) + 1$, we get $\min(S_w[i-1]) = S_w[i, j] - 2$ with $B_w[i-1, j] = 1$ the following holds:

$$\begin{aligned} S_w[i-1, j] - S_w[i, j] + 2 &= 1, \\ S_w[i, j] - S_w[i-1, j] &= 1, \\ |w[j, i+j-1]|_1 - |w[j, i+j-2]|_1 &= 1. \end{aligned}$$

Therefore, $w[i+j-1] = 1$ and $j \in D_1 - (i-1)$.

□

Example 60 For the same rational number $3/7 \pmod{10}$, by calculating the decreasing values we get:

$$D_1 = \{4, 7, 10\}$$

$$\begin{aligned}
D_2 &= D_1 \cup (D_1 - (1)) = \{4, 7, 10\} \cup \{3, 6, 9\} = \{3, 4, 6, 7, 9, 10\} \\
D_3 &= D_2 \cup (D_1 - (2)) = \{3, 4, 6, 7, 9, 10\} \cup \{2, 5, 8\} = \{2, 3, 4, 5, 6, 7, 8, 9, 10\} \\
D_4 &= D_3 \cap (D_1 - (3)) = \{2, 3, 4, 5, 6, 7, 8, 9, 10\} \cap \{1, 4, 7\} = \{4, 7\} \\
D_5 &= D_4 \cup (D_1 - (4)) = \{4, 7\} \cup \{3, 6, 10\} = \{3, 4, 6, 7, 10\} \\
D_6 &= D_5 \cup (D_1 - (5)) = \{3, 4, 6, 7, 10\} \cup \{2, 5, 9\} = \{2, 3, 4, 5, 6, 7, 9, 10\} \\
D_7 &= D_6 \cap (D_1 - (6)) = \{2, 3, 4, 6, 7, 9, 10\} \cap \{1, 4, 9\} = \{4\} \\
D_8 &= D_7 \cup (D_1 - (7)) = \{4\} \cup \{3, 7, 10\} = \{3, 4, 7, 10\} \\
D_9 &= D_8 \cup (D_1 - (8)) = \{3, 4, 7, 10\} \cup \{2, 6, 9\} = \{2, 3, 4, 6, 7, 9, 10\} \\
D_{10} &= D_9 \cap (D_1 - (9)) = \{2, 3, 4, 6, 7, 9, 10\} \cap \{1, 5, 8\} = \{\}.
\end{aligned}$$

Example 61 Let the rational number $2/7 \pmod{9}$, by calculating the decreasing values we get:

$$\begin{aligned}
D_1 &= \{5, 9\} \\
D_2 &= D_1 \cup (D_1 - (1)) = \{5, 9\} \cup \{4, 8\} = \{4, 5, 8, 9\} \\
D_3 &= D_2 \cup (D_1 - (2)) = \{4, 5, 8, 9\} \cup \{3, 7\} = \{3, 4, 5, 7, 8, 9\} \\
D_4 &= D_3 \cup (D_1 - (3)) = \{3, 4, 5, 7, 8, 9\} \cup \{2, 6\} = \{2, 3, 4, 5, 6, 7, 8, 9\} \\
D_5 &= D_4 \cap (D_1 - (4)) = \{2, 3, 4, 5, 6, 7, 8, 9\} \cap \{1, 5\} = \{5\} \\
D_6 &= D_5 \cup (D_1 - (5)) = \{5\} \cup \{4, 9\} = \{4, 5, 9\} \\
D_7 &= D_6 \cup (D_1 - (6)) = \{4, 5, 9\} \cup \{3, 8\} = \{3, 4, 5, 8, 9\} \\
D_8 &= D_7 \cup (D_1 - (7)) = \{3, 4, 5, 8, 9\} \cup \{2, 7\} = \{2, 3, 4, 5, 7, 8, 9\} \\
D_9 &= D_8 \cap (D_1 - (8)) = \{2, 3, 4, 5, 7, 8, 9\} \cap \{1, 6\} = \{\}.
\end{aligned}$$

The balance test matrix $B_{\frac{2}{7}}$ is given by:

$$B_{\frac{2}{7}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

Lemma 2.41 For the balance matrix $B_{\frac{a}{n}}$, we have: $\text{card}(D_i) = i.a \pmod{n}$.

Proof. The set D_1 contains exactly a elements, then $\text{card}(D_1) = 1.a = a$. By induction, we let $\text{card}(D_{i-1}) = (i-1).a \pmod{n}$ and we prove it true for D_i . From the previous lemma, we have if $i \notin D_1$, $D_i = D_{i-1} \cup (D_1 - (i-1))$ and since the two sets are distinct, we get:

$$\text{card}(D_i) = \text{card}(D_{i-1}) + \text{card}(D_1 - (i-1)) = (i-1)a \pmod{n} + a = i.a \pmod{n}.$$

When $i \in D_1$, we get:

$\text{card}(D_i) = \text{card}(D_{i-1}) + \text{card}(D_1 - (i-1)) - \text{card}(D_{i-1} \cup (D_1 - (i-1))) = (i-1)a \bmod n + a - n = ia \bmod n$. Since the union covers all the positions in the Christoffel word. \square

Note: This cardinality confirms that the period row is made only of zeros since $\text{card}(D_n) = n.a \bmod n = 0$, hence we have no values of 1 on this row.

Due to the symmetry in the Stern-Brocot tree and the fact that $E(C(\frac{a}{b})) = C(\frac{b}{a})$, we deduce the final lemma of this Chapter that allows us to reduce our study to the first half of the Stern-Brocot tree where $a < b$.

Lemma 2.42 *The balance matrix $B_{\frac{a}{b}}$ and $B_{\frac{b}{a}}$ are conjugate in the sense that:*

$$B_{\frac{a}{b}}[i] \equiv_{\alpha} E(B_{\frac{b}{a}}[i]) \quad \forall 1 \leq i < n; \text{ where } \alpha a \equiv -1 \bmod n.$$

Proof. Let a/b be a rational number with $n = a + b$, the Christoffel words $C(\frac{a}{b})$ and $C(\frac{b}{a})$ are conjugate and we have: $D(a, b) = \{(i\alpha \bmod n) + 1; i = 1 \dots a\}$ and $D(b, a) = \{(i\alpha' \bmod n) + 1; i = 1 \dots b\}$. From [38] in lemma 3.8; we have $C(\frac{a}{b}) \equiv_{\alpha} E(C(\frac{b}{a}))$; where $a\alpha \equiv_n -1$. Thus, $B_{\frac{a}{b}}[1] \equiv_{\alpha} E(B_{\frac{b}{a}}[1])$. The pattern is preserved for the other rows and the same relation holds since they are deduced from the first one. \square

Till here, we had an overview over all the notions and tools needed to reach our result. In the following Chapter, we introduce new tools and techniques and we explain their aim in studying the synchronization of Christoffel words.

Chapter 3

Second order balanced matrix

In the first chapter, we introduced several families of words and we mentioned that in this thesis our study is focusing on Christoffel words, which means that we restricted the work on a binary alphabet that is $A = \{0, 1\}$. This family of words is very rich with all its properties that are considered as strong tools in different domain of research. In addition to that, we gave the definition of the balancedness of a word w over an alphabet A . With a matrix called the balanced matrix B_w , we were able to provide in an explicit way the order of balancedness of a circular word. The initial definition is given for a binary word but later, a general form is provided for any circular word over a d alphabet. Moreover, and since we know that the Christoffel words are finite 1-balanced words as we have seen in the first chapter, therefore the elements of its balanced matrix are composed only of 0 and 1. This is equivalent to say that each row of this matrix is composed of a binary word. Here comes the advanced idea of studying a second order of balancedness for the Christoffel words. Therefore, in this chapter, we define another new matrix called the second order balance matrix, denoted U_w , that provides us a second order of balancedness for Christoffel words. This matrix is obtained in a recursive way over the rational numbers, which will be explained in detail in this chapter. This matrix is totally new, and with all its properties, will be considered as an useful tool in order to reach the synchronization of the Christoffel words.

In Figure 3.1, we see the Christoffel word of slope $3/5$, represented on the Cayley graph at level $k = 1$ with the respected positions of letters and the axis of symmetry as explained in Chapter 2. As we know, this word is 1-balanced since for any two factors of same length, the difference of number of occurrences for the letter 1 does not exceed 1. In the same figure, we can identify on level 2, the word $w' = 01101111$ which is the number of occurrences of the letter 1 in a sliding window of length 2. Strictly speaking, this level is the illustration of $B_{\frac{3}{5}}[2]$ which is a binary word. While for the third level, we get the word $w'' = 11111211$ that corresponds to a

sliding window of length 3, and by subtracting with the minimal value, we get $w'' = 00000100$ which is $B_{\frac{3}{5}}[3]$ that again is a binary word. And so on, we repeat this concept until level $n = 8$. Since each level is a binary word we can approach a new order of balancedness for the Christoffel words and test how the balance is affected in each level of this Cayley graph. In fact, this is what we will see in this Chapter which is the considerable part of this thesis. A second order of balancedness is defined and lot of properties and open questions arise due to this study. We will see all those details in the coming chapters while in this one an arithmetical and technical proof will be given to explain the construction of the matrix that studies this new order of balancedness.

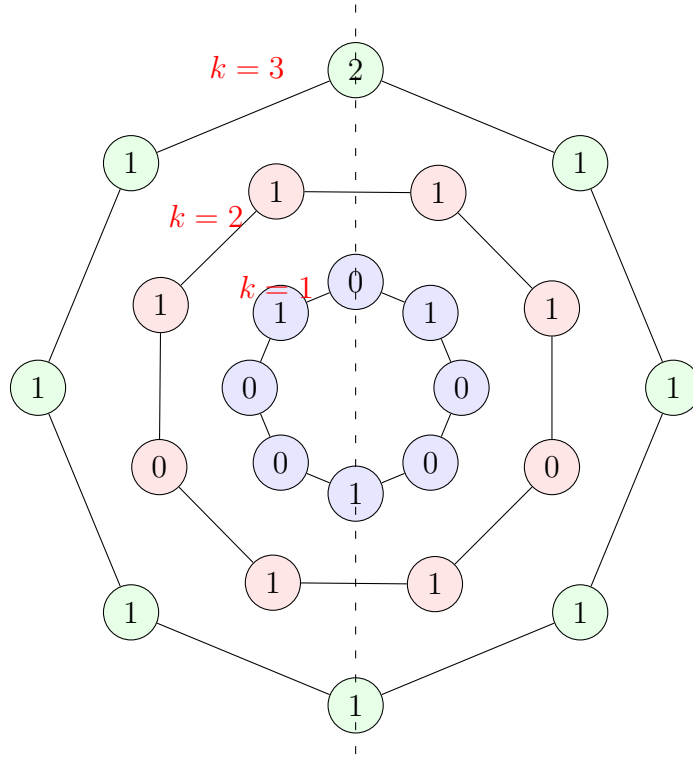


Figure 3.1: The first three levels of balancedness for the Christoffel word of slope $3/5$ representation on a cayley graph.

3.1 Second order balance matrix

We now investigate a refinement of balancedness in the case of 1-balanced circular words by computing two times the balance property. In fact, each row is a circular

binary word in the matrix B_w , therefore we are studying the balancedness of each row of B_w .

Definition 3.1 *Let w be a word such that w^ω is 1-balanced or, equivalently, that B_w is a binary matrix. The Second order balance matrix $U_w = (u_{ij})_{1 \leq i, j \leq n-1}$ is defined as:*

$$U_w[i, j] = \max(B_{(B_w[i])}[j]).$$

In other words, for a pair of integers i, j , consider the word $B_w[i]$ and list all its factors of length j . Among these factors, choose p a factor that maximizes the number of occurrences of the letter 1 and q a factor that minimizes it. The entry $U_w[i, j]$ is given by $|p|_1 - |q|_1$. Equivalently,

$$U_w[i, j] = \max_{v \in L_j(B_w[i])} |v|_1 - \min_{v \in L_j(B_w[i])} |v|_1.$$

Definition 3.2 *The second order of balance of a circular 1-balanced word w is:*

$$\delta^2(w) = \max(U_w).$$

Once again, in order to enlighten the notations, when working with the Christoffel $C(\frac{a}{b})$, the second order balance matrix $U_{C(\frac{a}{b})}$ is simply denoted $U_{\frac{a}{b}}$.

We recall that w denotes always a Christoffel word of slope a/b such that: $w = C(\frac{a}{b})$ of length $n = |w|$. The Second order balance matrix of a Christoffel word of slope $\frac{a}{b}$, $U_{\frac{a}{b}}$, is of dimension $(n-1) \times (n-1)$.

Example 62 *Let $\frac{a}{b} = \frac{3}{7}$ with $n = 10$. The balance matrix $B_{\frac{3}{7}}$ is:*

$$B_{\frac{3}{7}} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \end{pmatrix},$$

For example, let us choose $B_{\frac{3}{7}}[5] = [0, 0, 1, 1, 0, 1, 1, 0, 0, 1]$, and by computing the balance matrix for this word and checking the fifth row by taking all the blocks of length 5 of $B_{\frac{3}{7}}[5]$, we get the element $U_{\frac{3}{7}}[5, 5]$ of the second order balance of the Christoffel word $C(\frac{3}{7})$. This value is obtained from the difference between the two

windows:11011 and 00100 that appear in the binary word $\overline{00 \mid 11011 \mid 001}$. Hence $u_{5,5} = 4 - 1 = 3$, since they are the factors of length 5 that contains the maximum and the minimum number of occurrences of the letter 1. With the same procedure we fill all the entries of the second order balance matrix of $\frac{3}{7}$ represented below:

$$U_{\frac{3}{7}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 & 3 & 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix},$$

we denote: $\delta^2(C(\frac{3}{7})) = 3$.

Sage code The code used to obtain the Second order balance Matrix U_w where s denotes a Christoffel word.

```

1
2 def U (a,b):
3     n=a+b
4     return matrix( n-1, n-1, lambda i,j : max(Balance_Matrix(
5         Balance_Matrix(s)[i])[j]) )

```

Listing 3.1: Code of the Second order balance Matrix.

3.2 Recursive construction of the second order balance matrix

The Stern-Brocot tree contains all the irreducible fractions that are distributed all over the tree using the Farey sum. The fraction $1/1$ belongs to the first level ($k = 1$), whose its left (resp. right) ancestor is $0/1$ (resp. $1/0$). Using these three fractions and applying the Farey sum we get all the other fractions on this tree. The second order balance matrix of a rational number a/b studies the second order balance property on the Christoffel word of slope a/b . The Christoffel tree and Stern-Brocot tree are isomorphic, hence we can assume that the construction of $U_{\frac{a}{b}}$ must follow a certain pattern. At the beginning of this section, we give

some necessary properties for the matrix $U_{\frac{a}{b}}$ needed for the proof of the recursive construction. In the second part we give the general form of $U_{\frac{a}{b}}$.

We recall that the Stern-Brocot tree as defined in Chapter 2 is divided into two symmetrical parts, where on the left side, we have all the irreducible fractions of the form a/b with $a < b$ and on the right side, all the fractions of the form b/a . In the next part we prove that $U_{\frac{a}{b}} = U_{\frac{b}{a}}$, which restrict our study to the first left side of the tree.

3.2.1 Properties of the matrix $U_{\frac{a}{b}}$

From section 1.5.2, we have that the rows of the upper half of $B_{\frac{a}{b}}$ are complementary to the rows of the lower half. This induces symmetries in the matrix $U_{\frac{a}{b}}$. More precisely, the second order balance matrix $U_{\frac{a}{b}}$ of dimension $(n-1) \times (n-1)$ has horizontal, vertical and diagonal symmetries. The axes of symmetry are at position $\frac{n}{2}$ or between $\frac{n-1}{2}$ and $\frac{n+1}{2}$ depending on the parity of n .

Property 3.3 For any position (i, j) ,

$$U_{\frac{a}{b}}[i, j] = U_{\frac{a}{b}}[n-i, j] = U_{\frac{a}{b}}[i, n-j] = U_{\frac{a}{b}}[n-i, n-j].$$

Proof. By Lemme 1.26, we have that for any i , $1 \leq i \leq n-1$, $B_w[i] \equiv_{n-i} E(B_w[n-i])$ where $n = a + b$. Consequently :

$$\begin{aligned} U_w[i, j] &= \max(B_{(B_{\frac{a}{b}}[i])}[j]) && \text{(def.)} \\ &= \max(B_{(B_{\frac{a}{b}}[n-i])}[j]) && \text{(autocomplementarity of } B_{\frac{a}{b}}) \\ &= \max(B_{(B_{\frac{a}{b}}[n-i])}[n-j]) && \text{(autocomplementarity of } B_{B_{\frac{a}{b}}[n-i]}) \\ &= \max(B_{(B_{\frac{a}{b}}[i])}[n-j]). \end{aligned}$$

□

Moreover, $U_{\frac{a}{b}}$ has an extra diagonal symmetry;

Property 3.4 For the Christoffel word of slope a/b , $w = C(\frac{a}{b})$, the matrix U_w has a diagonal symmetry so that $(U_w)^T = U_w$.

Proof. We let $i < j$, by definition: $U_w[i, j] = \max(B_{(B_w[i])}[j])$; we recall that: $B_w[i] = S_w[i] - \min(S_w[i])$ where $S_w[i, j] = |s_w[i, j]|_1$; and $s_w[i, j] = w^\omega[j \dots j+i-1]$. Therefore $B_{B_w[i]}[j] = S_{B_w[i]}[j] - \min(S_{B_w[i]}[j])$ with $S_{B_w[i]}[j] = |s_{s_w[i, j]}[j, i]|_1 = |w[j, \dots, j+i-1](j, i)|_1 = |w[j, \dots, j+i-1]|_1$. While for $U[j, i] = \max(B_{(B_w[j])}[i])$

we have: $B_w[j] = S_w[j] - \min(S_w[j])$ where $S_w[j, i] = |s_w[j, i]|_1$; and $s_w[j, i] = w[i \dots j + i - 1]$. Therefore $B_{B_w[j]}[i] = S_{B_w[j]}[i] - \min(S_{B_w[j]}[i])$ with $S_{S_w[j]}[i] = |s_{s_w[j, i]}[i, j]|_1 = |w[i, \dots, j + i - 1][i, j]|_1 = |w[j, \dots, j + i - 1]|_1$. Then:

$$U_w[i, j] = \max(B_{(B_w[i])}[j]) = \max(B_{(B_w[j])}[i]) = U[j, i], \forall i, j \in \{1, \dots, n-1\}.$$

□

Property 3.5 Let $\frac{a}{b}$ be a rational number with $n = a + b$, we have: $U_{\frac{a}{b}} = U_{\frac{b}{a}}$.

Proof. From lemma 2.42, we have $B_{\frac{a}{b}}[i] \equiv_{\alpha} E(B_{\frac{b}{a}})[i], \forall i \in \{1, \dots, n\}$. Therefore,

$$U_{\frac{a}{b}}[i, j] = \max(B_{(B_{\frac{a}{b}}[i])}[j]) = \max(B_{(B_{\frac{b}{a}}[i])}[j]) = U_{\frac{b}{a}}[i, j].$$

□

Example 63 Let $w = C(\frac{5}{8})$, by evaluating the second order balance matrix $U_{\frac{5}{8}}$, we can observe the three axis of symmetries in this matrix.

$$U_{\frac{5}{8}} = \left(\begin{array}{cccccc|cccccc} \color{red}{1} & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \color{red}{1} & 1 & 2 & 1 & 1 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & \color{red}{1} & 1 & 1 & 2 & 2 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & \color{red}{2} & 1 & 2 & 2 & 1 & 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & \color{red}{1} & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 & 1 & \color{red}{2} & 2 & 1 & 2 & 2 & 1 & 1 \\ \hline 1 & 1 & 2 & 2 & 1 & 2 & \color{red}{2} & 1 & 2 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & \color{red}{1} & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 & 1 & 2 & 2 & 1 & \color{red}{2} & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 2 & 2 & 1 & 1 & \color{red}{1} & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 & 1 & 1 & 2 & 1 & \color{red}{1} & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \color{red}{1} \end{array} \right).$$

3.2.2 General form of the second order balance matrix

The second order balance matrix is obtained from the balance matrix that has lot of properties and a symmetry as shown in Chapter 1 and 2. Due to those properties, we were able to show that the matrix U_w has three axis of symmetry and that $U_{\frac{a}{b}} = U_{\frac{b}{a}}$. In addition to the symmetries, this matrix has a specific structure. In fact, it is divided into 9 blocks and due to the symmetries defined in

properties 3.3 and 3.4, it is sufficient to know three of them to deduce the others. But what are those blocks? Where are they located? How can we obtain them? Why are they located at that specific position? Many and many questions can be asked. In this section, we give the general form of the second order balance matrix and we give a complete construction with a detailed proof and interpretation for each block of this matrix.

We remark that in this matrix and for each Christoffel word, there are 2 rows and 2 columns full of 1 called **Synchronized rows**. Furthermore, we will understand the reason of having on these rows and columns the value 1. But in fact, to construct $U_{\frac{a}{b}}$, we start by placing the 4 lines of separation that divide the matrix into 9 blocks. Normally those lines are placed just after the first row and column containing only 1's and before the second row and column of the synchronized rows. As we said and due to the symmetries found in this matrix, we reduce the study to only three blocks which are enough to complete the full construction of $U_{\frac{a}{b}}$. These 3 blocks are noted α, β and γ and are located in the matrix in this way:

$$U_{\frac{a}{b}} = \left(\begin{array}{c|c|c} \alpha & \cdot 1 & \cdot 2 \\ \hline \gamma & \beta & \cdot 3 \\ \hline \cdot 4 & \cdot 5 & \cdot 6 \end{array} \right),$$

The blocks 1, 2, 3, 4, 5 and 6 are obtained from the three blocks by the following way:

1. Block 1 is the transpose of γ ;
2. Blocks 2, 4 and 6 are exactly the bloc α obtained by horizontal and vertical symmetries;
3. Block 3 is the block γ by vertical symmetry and block 5 is γ^T by horizontal symmetry.

In Chapter 2, we introduced some particular fractions in Definition 2.28. These fractions are very important for the construction of the second order balance matrix. The Christoffel word of slope a/b is the concatenation of some Christoffel words of smaller slopes. Therefore in the following section, we show that the blocks α, β, γ are described by the second order balance matrices of simpler fractions. The construction of the matrix $U_{\frac{a}{b}}$ depends on the position of the rational number on the Stern-Brocot tree. In fact the rational number can be either after the first deviation position i.e the last component of its continued fraction is equal to 2 or it can be on an extended branch which reveals that its last partial denominator is greater or equal to three. For both cases, the block α is deduced from $U_{\text{TBF}(\frac{a}{b})}$ while β is given by adding one to each entry of the second order balance matrix

of the 2-nd unidirectional father of $\frac{a}{b}$. While the construction of γ depends on the fact that in the Stern-Brocot tree, the fraction $\frac{a}{b}$ is, relatively to its father, a deviation (first deviation fraction) or an extension (first extended fraction). For this block and as a particular case, we will see in the next section the accurate steps needed to construct this block.

3.3 The construction of $U_{\frac{a}{b}}$

In this section, we define the tools used to construct recursively the second order balance matrix of the Christoffel word of slope $a/b = [a_0, \dots, a_z]$. Figure 3.3, displays some rational numbers that will be used later on. Since $U_{\frac{a}{b}} = U_{\frac{b}{a}}$, we reduce the work to the first half of the Stern-Brocot tree, that is the half containing all the fractions such with $a \leq b$. Given $\frac{a}{b} = [a_0, \dots, a_z]$ and since $a < b$ then $a_0 = 0$, we consider separately the cases $z < 2$ and $z \geq 2$.

3.3.1 The trivial cases : $z \in \{0, 1\}$

We start by the trivial cases for $z \in \{0, 1\}$:

- For $z = 0$, we have $U_{\frac{0}{1}} = U_{\frac{1}{0}} = []$ and $U_{\frac{1}{1}} = [1]$.
- Since we mentioned that the work is restrained to the rational numbers with $a < b$, then for $z = 1$, we have $\frac{a}{b} = [0, a_1]$ which implies that $a = 1$ and $a_1 = b$. Those fractions are located on the first extended branch on the Stern-Brocot tree.

From Remark 1.28, we have that $B_w[i] = 0^{b+1-i}1^i$ and Definition 3.1 states that: $U_w[i, j] = \max(B_{B_w[i]}[j])$. With these information, we can get a particular form for U_w where w is a Christoffel word of slope $1/b$.

Using Properties 3.3 and 3.4, we know that it is enough to determine the values of the first quarter of the matrix U_w and since we have a diagonal symmetry it suffices to consider $j \leq i$ where $i, j \leq \frac{n}{2}$ if n is even or $i, j \leq \frac{n-1}{2}$ if n is odd.

Proposition 3.6 *If $a = 1$, then for all $j \leq i$ and $i, j \leq \frac{n}{2}$ we have: $U_{\frac{a}{b}}[i, j] = j$.*

Proof. Since $B_w[i] = 0^{b+1-i}1^i$ and $U_w[i, j] = \max(B_{B_w[i]}[j])$, from Property 1.29 and with $j \leq i$, we have: $B_{B_w[i]}[j] = 0^{b-i-j+2}1 \dots j^{i-j+1} \dots 1$. Therefore, $U_w[i, j] = \max(B_{B_w[i]}[j]) = j$. \square

Example 64 Let $a/b = 1/8$, the second order balance matrix $U_{\frac{1}{8}}$ is given as follows:

$$U_{\frac{1}{8}} = \left(\begin{array}{cccc|cccc} \color{red}{1} & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \color{red}{1} & \color{red}{2} & 2 & 2 & 2 & 2 & 2 & 1 \\ \color{red}{1} & \color{red}{2} & \color{red}{3} & 3 & 3 & 3 & 2 & 1 \\ \color{red}{1} & \color{red}{2} & \color{red}{3} & \color{red}{4} & 4 & 3 & 2 & 1 \\ \hline 1 & 2 & 3 & 4 & 4 & 3 & 2 & 1 \\ 1 & 2 & 3 & 3 & 3 & 3 & 2 & 1 \\ 1 & 2 & 2 & 2 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \right).$$

We can note that this matrix can also be constructed using a recursive form where the first (resp. last) row and the first (resp. last) column are all 1's and in the middle we have the matrix of $SUF(\frac{1}{b})$ where its elements are increased by 1 (see Figure 3.2).

$$U_{\frac{1}{b}} = \left(\begin{array}{c|ccc|c} 1 & 1 & & \dots & 1 & 1 \\ \hline 1 & & & & & 1 \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ 1 & & & & & 1 \\ \hline 1 & 1 & & \dots & 1 & 1 \end{array} \right).$$

$U_{[0, a_1-2]} + 1$

Figure 3.2: The general form of the matrix $U_{\frac{1}{b}}$.

For the same example, we remark that:

$$U_{\frac{1}{6}} = \left(\begin{array}{cccccc|c} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 2 & 2 & 1 \\ 1 & 2 & 3 & 3 & 2 & 1 & \\ 1 & 2 & 3 & 3 & 2 & 1 & \\ 1 & 2 & 2 & 2 & 2 & 1 & \\ 1 & 1 & 1 & 1 & 1 & 1 & \end{array} \right) \quad \text{then} \quad U_{\frac{1}{8}} = \left(\begin{array}{cccccc|c} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 2 & 2 & 1 \\ 1 & 2 & 3 & 3 & 3 & 2 & 1 \\ 1 & 2 & 3 & 4 & 4 & 3 & 1 \\ 1 & 2 & 3 & 4 & 4 & 3 & 1 \\ 1 & 2 & 3 & 3 & 3 & 2 & 1 \\ 1 & 2 & 2 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \right).$$

3.3.2 The general case : $z \geq 2$

We now assume that $z \geq 2$, in order to enlighten the presentation we define the following fractions, let : $\frac{u}{v} = \text{TBF}(\frac{a}{b})$, $\frac{x}{y} = \text{FRF}(\frac{a}{b})$, $\frac{c}{d} = \text{FDF}(\frac{u}{v})$, $\frac{e}{f} = \text{FPF}(\frac{c}{d})$,

$\frac{g}{h} = \text{SUF}(\frac{u}{v})$, $\frac{p}{q} = \text{TBF}(\frac{x}{y})$, $\frac{s}{t} = \text{FRF}(\frac{u}{v})$. See Figure 3.3 for an illustration of their relative positions on the Stern-Brocot tree.

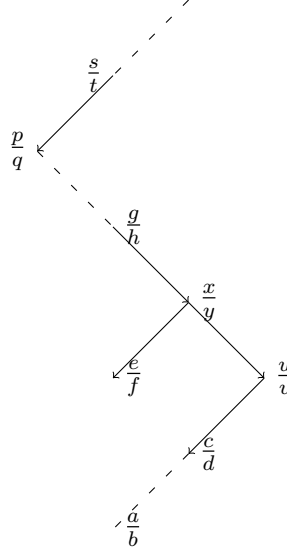


Figure 3.3: Position of the fractions $\frac{u}{v}$, $\frac{x}{y}$, $\frac{c}{d}$, $\frac{e}{f}$, $\frac{g}{h}$, $\frac{p}{q}$ and $\frac{s}{t}$ relatively to $\frac{a}{b}$, for the case where $a_{z-1} \geq 2$ and z odd. By definition, we have: $\frac{a}{b} = [a_0 \dots, a_z]$, $\frac{u}{v} = [a_0 \dots, a_{z-1} + 1]$, $\frac{x}{y} = [a_0, \dots, a_{z-1}]$, $\frac{p}{q} = [a_0, \dots, a_{z-2} + 1]$, $\frac{s}{t} = [a_0, \dots, a_{z-2}]$, $\frac{e}{f} = [a_0, \dots, a_{z-1} - 1, 2]$, $\frac{c}{d} = [a_0, \dots, a_{z-1}, 2]$, $\frac{g}{h} = [a_0, \dots, a_{z-1} - 1]$.

With respect to the Farey addition we have: $\frac{u}{v} = \frac{x}{y} \frac{s}{t}$ and $\frac{c}{d} = \frac{x}{y} \frac{u}{v}$.

Lemma 3.7 *The Christoffel words of slope $\frac{a}{b}$, $\frac{g}{h}$ and $\frac{e}{f}$ are :*

- a) $C(\frac{a}{b}) = \left(C(\frac{x}{y})\right)^{a_z-1} C(\frac{u}{v}) = \left(C(\frac{x}{y})\right)^{a_z} C(\frac{s}{t})$.
- b) $C(\frac{x}{y}) = C(\frac{g}{h})C(\frac{s}{t})$.
- c) $C(\frac{e}{f}) = C(\frac{g}{h})C(\frac{x}{y})$.

Proof. This is a consequence of Theorem 1.10 and Remark 2.29. The equalities are obtained by iteration of the functions ϕ_0 and ϕ_1 . In fact, by construction and using the Farey sum, $\frac{a}{b} = \left(\frac{x}{y}\right)^{a_z-1} \left(\frac{u}{v}\right)$. Due to the isomorphism between the Christoffel tree and stern-Brocot tree, we get the same relations between the Christoffel words. Same reasoning is used for the other relations. \square

Separation lines

For the rest of this section, the fractions mentioned in Figure 3.3 are used to prove the construction of $U_{\frac{a}{b}}$. Now we prove that this matrix is decomposed into 9 blocks. In fact, the first step is to place the 4 separation lines in the right place in order to get the 9 blocks of the matrix. As we mentioned in section 3.2.2, those separation lines depend on the place of the rows and columns of 1 in the matrix. Therefore, we show in Lemma 3.8, the exact position of these synchronized rows. We recall that $\frac{x}{y} = \text{FRF}(\frac{a}{b}) = [0, \dots, a_{z-1}]$, and we let $k = x + y$.

Lemma 3.8 *Let $k = x + y$, for the rational number $\frac{a}{b}$ with $n = a + b$, we have:*

$$U[k, j] = U[n - k, j] = 1, \quad \forall j \in \{1, \dots, n - 1\}$$

and

$$U[i, k] = U[i, n - k] = 1, \quad \forall i \in \{1, \dots, n - 1\}.$$

Proof. From Properties 2.17 and 2.16, we have $(x + y)a = x(a + b) - 1$, and using the consecutive fractions a/b and x/y from the Stern-Brocot tree, we get: $ka = xn - 1$ hence $ka \equiv -1 \pmod{n}$. From lemma 2.41, we have $\text{card}(D_i) = ia \pmod{n}$, hence $\text{card}(D_k) = n - 1$ and $\text{card}(D_{n-k}) = 1$ since these two rows are complementary. Therefore, the words $B_{\frac{a}{b}}[k]$ and $B_{\frac{a}{b}}[n - k]$ are 1-balanced; hence $U[k, j] = U[n - k, j] = 1, \forall j \in \{1, \dots, n - 1\}$. Using the diagonal symmetry of $U_{\frac{a}{b}}$, we get: $U[i, k] = U[i, n - k] = 1, \forall i \in \{1, \dots, n - 1\}$. \square

Remark 3.9 *From Lemma 3.8, we get that the row (respectively column) k and $n - k$ contains only values of 1's. Therefore, we define the separation lines to be between the rows (resp. columns) $(k; k + 1)$ and between the rows (resp. columns) $(n - k - 1; n - k)$ which divide $U_{\frac{a}{b}}$ into 9 blocks. See Figure 3.4.*

Initial blocks

In this part, we prove that each block α, β and γ of $U_{\frac{a}{b}}$ is formed from some smaller second order balanced matrices.

To construct the second order balance matrix $U_{\frac{a}{b}}$ we first place the lines of separation. This shows that $U_{\frac{a}{b}}$ is composed of 9 blocks where the three blocks $\alpha_{k \times k}$, $\beta_{(n-2k-1) \times (n-2k-1)}$ and $\gamma_{(n-2k-1) \times k}$ are constructed while the others are obtained by symmetry.

In the following part, we construct respectively each of the blocks α, γ and β to get $U_{\frac{a}{b}}$.

Let $\frac{a}{b} = \frac{5}{8}$, the 9 blocks are represented as follows:

$$U_{\frac{5}{8}} = \begin{pmatrix} \begin{array}{ccccc|cc|cc|cc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 2 & 2 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 & 1 & 2 & 2 & 1 & 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 2 & 2 & 1 & 2 & 2 & 1 & 2 & 2 & 1 & 1 \\ 1 & 1 & 2 & 2 & 1 & 2 & 2 & 1 & 2 & 2 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 2 & 1 & 2 & 1 & 2 & 2 & 1 & 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 2 & 2 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \end{pmatrix}$$

Top left part of $U_{\frac{3}{5}}$ ←

→ $U_{\frac{1}{2}} + 1$

Left central part of $U_{\frac{3}{4}}$ ←

Figure 3.4: The decomposition of the matrix $U_{\frac{5}{8}}$ into 9 blocks. We have $\text{TBF}(\frac{5}{8}) = \frac{3}{5}$, $\text{SUF}(\frac{5}{8}) = \frac{1}{2}$, $\text{FPF}(\frac{5}{8}) = \frac{3}{4}$, $\frac{x}{y} = \frac{2}{3}$ and $k = 5$.

α -block:

Recall that $z \geq 2$ and we let for the rest of this section z to be odd where the second case is obtained in a similar way. Let $\frac{s}{t}$, $\frac{g}{h}$ and $\frac{x}{y}$ be the fractions defined at the beginning of Section 3.3.2. The α -block is formed of k rows and columns where $k = x + y = |\text{FRF}(\frac{a}{b})|$.

Lemma 3.10 *The α -block of $U_{\frac{a}{b}}$ is exactly the first $k - 1$ rows and columns of the matrix $U_{\frac{u}{v}}$ where $\frac{u}{v} = \text{TBF}(\frac{a}{b})$.*

Proof. First we show that the $k - 1$ first rows of the matrix $B_{\frac{a}{b}}$ are exactly the concatenation between the matrix $B_{\frac{x}{y}}$ repeated $a_z - 1$ times and the $k - 1$ first rows of the matrix $B_{\frac{u}{v}}$. Then we show that the language of the words obtained in the first $k - 1$ rows at position j , for $1 \leq j \leq k - 1$, is exactly the same as the language of the words obtained in the first $k - 1$ rows of $B_{\frac{u}{v}}$.

By Lemma 3.7, we have: $C(\frac{a}{b}) = C(\frac{x}{y})^{a_z-1}C(\frac{u}{v})$ and $C(\frac{u}{v}) = C(\frac{x}{y})C(\frac{s}{t})$ since z is odd. To construct $B_{\frac{a}{b}}$, we need to consider all the sliding windows of length between 1 and $n - 1$, then test the occurrences of the letter 1 for the circular Christoffel word of slope a/b at all the positions from 1 to n . Therefore, $B_{\frac{a}{b}}[i, j]$ for $1 \leq i, j \leq k - 1$ is exactly the same as $B_{\frac{x}{y}}[i, j]$, since all the sliding windows of different length can not exceed the k^{th} position. This means that all these factors are included in the prefix of length $2k$ of $C(\frac{a}{b})$ which is exactly $C(\frac{x}{y})^2$.

The same argument holds for each $B_{\frac{a}{b}}[i, j]$, for $1 \leq i \leq k - 1$ and $1 \leq j \leq n - (u + v)$ which form the $a_z - 1$ blocks composed of $B_{\frac{x}{y}}$. Knowing that the last block is also obtained because we know that $C\left(\frac{u}{v}\right)$ has $C\left(\frac{x}{y}\right)$ as a prefix. For the entries of $B_{\frac{a}{b}}[i, j]$ where $1 \leq i \leq k - 1$ and $n - (u + v) + 1 \leq j \leq n$, we have exactly the same values as in $B_{\frac{u}{v}}$ since, again, we only consider factors up to length k and both words $C\left(\frac{a}{b}\right)$ and $C\left(\frac{u}{v}\right)$ start with $C\left(\frac{x}{y}\right)$ which is of length k .

Hence we have: $B_{\frac{a}{b}}[i] = B_{\frac{x}{y}}[i]^{a_z - 1} \cdot B_{\frac{u}{v}}[i]$ for all $1 \leq i \leq k - 1$.

Once this relation is set, we can check that $L_j(C\left(\frac{x}{y}\right)^\omega) \subseteq L_j(C\left(\frac{u}{v}\right)^\omega)$ for $j \leq k - 1$ since $C\left(\frac{x}{y}\right)$ ends with $C\left(\frac{s}{t}\right)$ and we have $C\left(\frac{u}{v}\right) = C\left(\frac{x}{y}\right)C\left(\frac{s}{t}\right)$. This implies

$$L_j\left(B_{\frac{a}{b}}^\omega[i]\right) = L_j\left(B_{\frac{u}{v}}^\omega[i]\right) \quad \forall 1 \leq i \leq k - 1 \quad \text{and} \quad 1 \leq j \leq k - 1.$$

□

γ -block:

The second block of $U_{\frac{a}{b}}$ is the γ -block of dimension $(n - 2k - 1) \times k$. This block is located between the $k + 1$ and the $(n - k - 1)^{th}$ rows and bounded by the line of separation at the k^{th} column. This part of the second order balance matrix admits a recursive form. If we consider the concatenation of matrices as a vertical stacking, then $\gamma\left(\frac{a}{b}\right)$ is given by:

$$\gamma\left(\frac{a}{b}\right) = \begin{cases} \gamma\left(\frac{c}{d}\right) \left[\gamma\left(\frac{u}{v}\right) \gamma\left(\frac{c}{d}\right) \right]^{a_z - 2} & \text{if } a_z \geq 3 \\ \gamma\left(\frac{e}{f}\right) & \text{if } a_z = 2 \end{cases} \quad (3.1)$$

The following lemma states this property in a formal way.

Lemma 3.11 *The γ -block is obtained depending on the position of the fraction in the Stern-Brocot tree.*

- If $a_z \geq 3$ then $\gamma\left(\frac{a}{b}\right)$ is obtained by stacking vertically $\gamma\left(\frac{c}{d}\right)$ over $a_z - 2$ copies of $\gamma\left(\frac{u}{v}\right)$ and $\gamma\left(\frac{c}{d}\right)$.
- Otherwise, if $a_z = 2$, then $\gamma\left(\frac{a}{b}\right)$ is given the extension of $\gamma\left(\frac{e}{f}\right)$ to the k -th column of $U_{\frac{e}{f}}$.

where $\frac{u}{v} = \text{TBF}\left(\frac{a}{b}\right)$; $\frac{c}{d} = \text{FDF}\left(\frac{u}{v}\right)$; $\frac{e}{f} = \text{FPF}\left(\frac{a}{b}\right)$.

Proof. By the Theorems 2.8 and 2.40, we have that $D(a, b) = \{i\alpha + 1 \pmod n\}$ with $\frac{a}{b} = [a_0, \dots, a_z]$ and since z is odd then by Property 2.38 we have $\alpha = n - k$. Let $\frac{x}{y} = [a_0, \dots, a_{z-1}]$, we recall that $|C(\frac{x}{y})| = x + y = k$, and $|C(\frac{s}{t})| = s + t = k''$. Let p_1, p_2 be the two central words of $C(\frac{s}{t})$ and $C(\frac{g}{h})$ so that $C(\frac{x}{y}) = 0p_210p_11 = 0p_101p_21$ the last equality holds because the central word p_210p_1 is a palindrome.

- **Case A:** If $a_z \geq 3$.

For $a_z = 3$, we have:

$$\begin{aligned} - C(\frac{a}{b}) &= \left(C(\frac{x}{y})\right)^2 C(\frac{u}{v}) = \left(C(\frac{x}{y})\right)^3 C(\frac{s}{t}); n = a + b. \\ - C(\frac{c}{d}) &= C(\frac{x}{y})C(\frac{u}{v}) = \left(C(\frac{x}{y})\right)^2 C(\frac{s}{t}); n' = c + d. \\ - C(\frac{u}{v}) &= C(\frac{x}{y})C(\frac{s}{t}); n'' = u + v. \\ - C(\frac{x}{y}) &= C(\frac{g}{h}).C(\frac{s}{t}) = 0p_210p_11. \end{aligned}$$

Using Theorem 2.8 and Property 2.38, we denote by D_1, D'_1, D''_1, X and S the sets of decreasing positions of the Christoffel words of slopes: $\frac{a}{b}, \frac{c}{d}, \frac{u}{v}, \frac{x}{y}$ and $\frac{s}{t}$ respectively defined as follows:

1. $D_1 = D(a, b) = \{i\alpha + 1 \pmod n; 1 \leq i \leq a\}$; where $\alpha = n - k$,
2. $D'_1 = D(c, d) = \{i\alpha' + 1 \pmod n'; 1 \leq i \leq c\}$; where $\alpha' = n' - k$,
3. $D''_1 = D(u, v) = \{i\alpha'' + 1 \pmod n''; 1 \leq i \leq u\}$; where $\alpha'' = k''$,
4. $X = D(x, y)$,
5. $S = D(s, t)$.

We remark that: $k'' = |p_1| + 2$ and $\alpha = 2k + k''$.

By construction, we have:

$$\begin{aligned} - X &= D(g, h) \cup \{S + (k - k'')\} \text{ since } C(\frac{x}{y})[k - k' + 1 \dots k] = C(\frac{s}{t}). \\ - C(\frac{u}{v}) &= C(\frac{x}{y})C(\frac{s}{t}), \text{ hence } D''_1 = X \cup \{S + k\}, \text{ where the two sets are} \\ &\text{disjoints and } \forall s \in S, s + k \leq n''. \end{aligned}$$

Similarly and for $k = \max(X)$ we get:

$$D'_1 = X \cup \{X + k\} \cup \{S + 2k\} \tag{3.2}$$

$$D_1 = X \cup \{X + k\} \cup \{X + 2k\} \cup \{S + 3k\}. \tag{3.3}$$

The following arithmetic relations exist between n, n' and n'' :

$$- n'' = k + k'' = \alpha',$$

- $n = 3k + k'' = 2k + n'' = 2k + \alpha' = \alpha + k = k + n'$,
- $n' = 2k + k'' = k + n'' = k + \alpha' = \alpha$,
- $n - n' = k = \alpha - \alpha' = n' - \alpha' = n' - n''$.

The $C(\frac{a}{b})$ is a 1-balanced word with $a < b$.

- For $i = a - 1$, we have $(a - 1)\alpha + 1 = a\alpha - \alpha + 1 = -\alpha \equiv_n k$, hence $k \in D_1 \Rightarrow k + 1 \notin D_1$.
- For $i = 1$, we get $\alpha + 1 \in D_1$. $\alpha + 1 \in D_1 \Rightarrow \alpha \notin D_1 \Rightarrow n' \notin D_1$, where $\alpha' = n' - 2k$, therefore $D'_1 \setminus \{n'\} \subseteq D_1$.

By lemma 2.39, the word $B_{\frac{a}{b}}[k] = 0^\alpha 10^{n-\alpha-1}$, hence $D_k = \{\alpha + 1\}$. Since $k + 1 \notin D_1$, by theorem 2.40, we get:

$$\begin{aligned} D_{k+1} &= D_k \cup \{D_1 - k\} \\ &= \{\alpha + 1\} \cup \{i\alpha + 1 - k \pmod n\} \end{aligned}$$

Since $i\alpha + 1 - k = i\alpha + 1 - n + \alpha \equiv_n (i + 1)\alpha + 1$ for all $1 \leq i \leq a$, we obtain:

- For $1 \leq i \leq a - 1$, all the elements of $D_1 \setminus \{\alpha + 1\}$ belong to D_{k+1} .
- For $i = a$, we get: $(a + 1)\alpha + 1 \equiv_n \alpha$, and hence α is an element of D_{k+1} .

Therefore, we conclude that: $D_{k+1} = D_1 \cup \{\alpha\}$. The same reasoning is repeated for $C(\frac{c}{d})$, and we get

$$D'_{k+1} = D'_1 \cup \{\alpha'\} = D'_1 \cup \{\alpha - n + n'\} = D'_1 \cup \{n''\}.$$

By construction, we have: $D'_1 \setminus \{n'\} \subset D_1 \subset D_{k+1}$ and $\{n'\} \subset D_{k+1}$, therefore $D'_1 \subset D_{k+1}$. Since $\alpha = 2k + k'' = 2|C(\frac{x}{y})| + |p_1| + 2$ and $D_{k+1} = D_1 \cup \{\alpha\}$, we have:

$$B_{\frac{a}{b}}[k+1] = C(\frac{x}{y})C(\frac{x}{y})0p_11p_21C(\frac{s}{t}) \quad \text{and} \quad B_{\frac{c}{d}}[k+1] = C(\frac{x}{y})0p_11p_21C(\frac{s}{t}).$$

Therefore: $B_{\frac{a}{b}}[k + 1][k + 1 \dots n] = B_{\frac{c}{d}}[k + 1]$. We have $|C(\frac{x}{y})| = k$ and, by abuse of notation, we consider the rows of $B_{\frac{x}{y}}$ in a circular way so that $B_{\frac{x}{y}}[k + 1] = B_{\frac{x}{y}}[1]$ which allows to write: $B_{\frac{a}{b}}[i] = B_{\frac{x}{y}}[i]B_{\frac{c}{d}}[i]$ where this relation holds $\forall k + 1 \leq i \leq \alpha' = n - 2k$.

Therefore, $L_j(B_{\frac{a}{b}}^\omega[i]) = L_j(B_{\frac{c}{d}}^\omega[i])$ for all $k + 1 \leq i \leq \alpha'$ and $1 \leq j \leq k$.

Hence we have $\forall k + 1 \leq i \leq \alpha'; 1 \leq j \leq k$:

$$U_{\frac{a}{b}}[i, j] = \max \left(B_{B_{\frac{a}{b}}[i]}[j] \right) = \max \left(B_{B_{\frac{a}{b}}[i]}[j] \right) = U_{\frac{a}{b}}[i, j].$$

Using Property 2.39, we have $B_{\frac{a}{b}}[2k] = B_{\frac{a}{b}}[n - \alpha'] = 0^{\alpha'} 10^{k-1} 10^{k-1}$. By Property 1.26, we get: $B_{\frac{a}{b}}[\alpha'] = 01^{k-1} 01^{\alpha-1}$, hence $D_{\alpha'} = \{2, \dots, k, k + 2, \dots, n\}$. Since $n - n' = n' - \alpha'$, we get $n = 2n' - \alpha' = 2\alpha - \alpha'$. For $i = 2$, we have $2\alpha + 1 = n + \alpha' + 1 \equiv_n \alpha' + 1$, which implies that $\alpha' + 1 \in D_1$ and $\alpha' \notin D_1$. Using Theorem 2.40 we get:

$$\begin{aligned} D_{\alpha'+1} &= D_{\alpha'} \cap \{D_1 - \alpha'\} \\ &= D_{\alpha'} \cap \{i\alpha + 1 - \alpha' \pmod n; \forall 1 \leq i \leq a\} \\ &= D_{\alpha'} \cap \{i\alpha + 1 - \alpha + k \pmod n; \forall 1 \leq i \leq a\} \\ &= D_{\alpha'} \cap \{(i - 2)\alpha + 1 \pmod n; \forall 1 \leq i \leq a\}. \end{aligned}$$

We have:

- for all $3 \leq i \leq a$, $D_1 \setminus \{k, n\} \subseteq D_{\alpha'+1}$.
- For $i = 1$, we get the position $(k + 1)$.
- For $i = 2$, we get the position 1.
- For $i = a + 1$ and $i = a + 2$, we get respectively the positions k , and n .

Therefore, $D_{\alpha'+1} = D_1 \setminus \{k, n\}$.

Since $\alpha'' = k''$, then: $B_{\frac{a}{b}}[k''] = 01^{n''-1} \Rightarrow D_{\alpha''} = \{2, \dots, n''\}$.
As $\alpha'' \notin D_1''$, hence $\alpha'' + 1 \in D_1''$ and by Theorem 2.40:

$$\begin{aligned} D_{\alpha''+1}'' &= D_{\alpha''}'' \cap \{D_1'' - \alpha''\} \\ &= \{2, \dots, n''\} \cap \{(i - 1)\alpha'' + 1 \pmod n''; \forall 1 \leq i \leq n''\}. \end{aligned}$$

1. For $i = u + 1$; we get $u\alpha'' + 1 = -1 + 1 = 0$,
2. For $i = 1$, we get the position 1,
3. For $2 \leq i \leq u$, we get: $D_1'' \setminus \{0\} = D_1'' \setminus \{n''\} = D_1'' \setminus \{\alpha'\}$.

Therefore, $D_{\alpha''+1}'' = D_1'' \setminus \{n''\} \subset D_1$ and then, $D_{\alpha''+1}'' \setminus \{k\} \subset D_{\alpha'+1}$.

We have: $B_{\frac{a}{b}}[\alpha' + 1] = 0p_2 10p_1 00p_2 10p_1 1C(\frac{x}{y})0p_1 0$ and $B_{\frac{a}{b}}[\alpha'' + 1] = C(\frac{x}{y})0p_1 0$.
Hence we get $B_{\frac{a}{b}}[\alpha' + 1][2k + 1, \dots, n] = B_{\frac{a}{b}}[\alpha'' + 1]$ and we can write:

$$B_{\frac{a}{b}}[\alpha' + 1] = 0p_2 10p_1 0C(\frac{x}{y})B_{\frac{a}{b}}[\alpha'' + 1],$$

where the relation remains true $\forall \alpha' + 1 \leq i \leq 2k$.

Therefore, $L_j \left(B_{\frac{a}{b}}^\omega[i] \right) = L_j \left(B_{\frac{u}{v}}^\omega[i] \right)$ for all $\alpha' + 1 \leq i \leq 2k$ and $1 \leq j \leq k$ since the element $0p_100p_21$ of length k appears in both languages thus we have $\forall \alpha' + 1 \leq i \leq 2k; 1 \leq j \leq k$:

$$U_{\frac{a}{b}}[i, j] = \max \left(B_{B_{\frac{a}{b}}[i]}[j] \right) = \max \left(B_{B_{\frac{u}{v}}[i]}[j] \right) = U_{\frac{u}{v}}[i, j].$$

The rows from $2k + 1$ till $n - k$ are obtained by symmetry since $2k > \frac{n}{2}$ and $U_{\frac{a}{b}}$ is symmetric. We get: $\gamma \left(\frac{a}{b} \right) = \gamma \left(\frac{c}{d} \right) \left[\gamma \left(\frac{u}{v} \right) \gamma \left(\frac{c}{d} \right) \right]^1$, where $1 = a_z - 2$.

When $a_z > 3$, we have $C \left(\frac{a}{b} \right) = \left(C \left(\frac{x}{y} \right) \right)^{a_z - 1} C \left(\frac{u}{v} \right)$, where $D_1 = X \cup \{X + k\} \cup \{X + 2k\} \dots \cup \{X + (a_z - 1)k\} \cup \{S + a_z k\}$. Using the same reasoning we get $\gamma \left(\frac{a}{b} \right) = \gamma \left(\frac{c}{d} \right) \left[\gamma \left(\frac{u}{v} \right) \gamma \left(\frac{c}{d} \right) \right]^{a_z - 2}$.

• **Case B:** If $a_z = 2$.

For this part we have: $C \left(\frac{a}{b} \right) = C \left(\frac{x}{y} \right) C \left(\frac{u}{v} \right)$ and we need the word $C \left(\frac{e}{f} \right) = C \left(\frac{g}{h} \right) C \left(\frac{x}{y} \right)$, where $\frac{e}{f} = \text{FPF} \left(\frac{a}{b} \right)$. The same notations are used as in the case A, where $n = a + b$, $n' = e + f$, $k = |C \left(\frac{x}{y} \right)|$ and $k' = |C \left(\frac{g}{h} \right)| = |p_2| + 2$. The sets $D_1 = D(a, b)$ and $D'_1 = D(e, f)$ are calculated with the same way and we get:

- $n = k + \alpha$,
- $n' = k + k' = k' + \alpha'$,

hence we get: $k = \alpha'$. Using the same technique, we get:

1. $D_{k+1} = D_1 \cup \{\alpha\}$;
2. $D'_{k'+1} = D'_1 \cup \{\alpha'\} = D'_1 \cup \{k\}$;
3. $k + 1 \notin D_1 \Rightarrow k + 1 \notin D_1 \cup \{\alpha\}$;
4. $\alpha' = k \in D_1$.

The rest of the positions are obtained due to the concatenation of the word. We have $D'_{k'+1} \setminus \{\alpha' + 1\} \subseteq D_{k+1}$; which means:

$$\{D'_1 \cup \{\alpha'\}\} \setminus \{\alpha' + 1\} \subseteq D_1 \cup \{\alpha\}.$$

The Christoffel word of slope $\frac{a}{b}$ can be written as:

$$\begin{aligned} C \left(\frac{a}{b} \right) &= 0p_210p_11.0p_210p_11.0p_11 \\ &= 0p_210p_11.0p_101p_21.0p_11 \end{aligned}$$

and since $\alpha = n - k = n - (|p_1| + |p_2| + 4)$, we get:

$$\begin{aligned} B_{\frac{a}{b}}[k+1] &= 0p_210p_11.0p_111p_21.0p_11 \\ &= 0p_101p_21.0p_111p_21.0p_11. \end{aligned}$$

While for $C(\frac{e}{f})$, we have:

$$\begin{aligned} C(\frac{e}{f}) &= 0p_21.0p_21.0p_11 \\ &= 0p_21.0p_101p_21 \end{aligned}$$

with $\alpha' = n' - k' = n' - (|p_2| + 2)$, hence $B_{\frac{e}{f}}[k'+1] = 0p_21.0p_111p_21 = 0p_101p_211p_21$. Therefore we get: $L_j(B_{\frac{a}{b}}^\omega[i]) = L_j(B_{\frac{e}{f}}^\omega[i])$ for all $k+1 \leq i \leq n-k-1$ and $1 \leq j \leq k$

thus $\forall k+1 \leq i \leq n-k-1; 1 \leq j \leq k$, we have:

$$U_{\frac{a}{b}}[i, j] = \max(B_{B_{\frac{a}{b}}[i]}[j]) = \max(B_{B_{\frac{e}{f}}[i]}[j]) = U_{\frac{e}{f}}[i, j].$$

□

β -block

The β -block is the center of $U_{\frac{a}{b}}$, the last block needed to complete the construction of the second order balance matrix. This block is of dimension $(n-2k-1) \times (n-2k-1)$ and located between the rows (respectively columns) k and $n-k$.

Lemma 3.12 *The β -block of $U_{\frac{a}{b}}$ is exactly the second order balance matrix of the fraction $\frac{a}{b} = \text{SUF}(\frac{a}{b})$ in the Stern-Brocot tree where its elements are increased by 1.*

Proof. By definition 3.1, $U_{\frac{a}{b}}$ studies the maximal unbalancedness of the Christoffel word of slope a/b where $u_{ij} = \max_{u \in L_j(B_w[i])} |u|_1 - \min_{u \in L_j(B_w[i])} |u|_1$. The elements of the γ -block are obtained for $k+1 \leq i \leq n-k-1$ and $1 \leq j \leq k$, which refer to the words $u \in L_j(B_w[i])$ where $B_w[i, j] = S_w[i, j] - \min(S_w[i])$ and $S_w[i, j] = |s_w[i, j]|_1 = |w^\omega[j, j+i-1]|_1$. The elements of the β -block are obtained using the words $u \in L_j(B_w[i])$ for $k+1 \leq i \leq n-k-1$ and $k+1 \leq j \leq n-k-1$. We have 2 cases:

- If $a_z \geq 3$: In the previous section we proved that for $a_z = 3$, we have:

- $B_{\frac{a}{b}}[k+1] = C(\frac{x}{y})C(\frac{x}{y})C(\frac{u}{v}) = C(\frac{x}{y})C(\frac{x}{y})0p_111p_21C(\frac{s}{t})$.
- $\text{SUF}(\frac{a}{b}) = \frac{u}{v}$ with $C(\frac{u}{v}) = 0p_210p_11.0p_21$.
- The factor $w' = 0p_111p_210 \in L_{k+1}(B_{\frac{a}{b}}^\omega[k+1])$.
- The factor $w'' = 0p_101p_210 \in L_{k+1}(B_{\frac{u}{v}}^\omega[1])$.

We remark that $|w'|_1 = |w''|_1 + 1$. Since $B_{\frac{a}{b}}[k+1] = C(\frac{x}{y})C(\frac{x}{y})w'p_21$ then by identifying the elements of $L_j(B_{\frac{a}{b}}^\omega[k+1])$, $\forall k+1 \leq j \leq n-k-1$, we remark that the factor that has the maximal occurrences of the letter 1 is the factor w' . By studying the following rows of $B_{\frac{a}{b}}$, the same factor causes the increase of the maximal value. Hence, $\forall k+1 \leq i \leq n-k-1$; $k+1 \leq j \leq n-k-1$:

$$U_{\frac{a}{b}}[i, j] = \max\left(B_{B_{\frac{a}{b}}[i]}[j]\right) = \max\left(B_{B_{\frac{u}{v}}[i-k]}[j-k]\right) + 1 = U_{\frac{u}{v}}[i-k, j-k] + 1.$$

While for $a_z > 3$, and due to the concatenation with the $C(\frac{x}{y})$ from the left side for each fraction, we always get that:

$$U_{\frac{a}{b}}[i, j] = U_{\frac{e}{\theta}}[i-k, j-k] + 1.$$

- If $a_z = 2$, we let $\frac{a}{b} = [a_0, \dots, 2]$ and we have from the previous sections:

- $B_{\frac{a}{b}}[k+1] = C(\frac{x}{y})C(\frac{u}{v}) = C(\frac{x}{y})0p_111p_21C(\frac{s}{t}) = 0p_210p_110p_111p_210p_11$.
- By Definition 2.2.1, $\text{SUF}(\frac{a}{b}) = \frac{s}{t} = [0, \dots, a_{z-2}]$, where $C(\frac{s}{t}) = 0p_11$.

Just like the previous case the factor that causes the difference is the factor $w' = 0p_111p_210$ that belongs to $L_{k+1}(B_{\frac{a}{b}}^\omega[k+1])$. All the other factors of this language will have the same occurrence of 1 except w' that will be increased by 1. Therefore, by taking the maximal occurrence of those factors in this language, the maximal value appears in $|w'|_1$ that gives the same value if we study the occurrence of 1 in all the factors of $L_1(B_{\frac{s}{t}}^\omega[1])$. By studying the following rows of $B_{\frac{a}{b}}$, the same factor causes the increase of the maximal value. Hence, $\forall k+1 \leq i \leq n-k-1$; $k+1 \leq j \leq n-k-1$:

$$U_{\frac{a}{b}}[i, j] = \max\left(B_{B_{\frac{a}{b}}[i]}[j]\right) = \max\left(B_{B_{\frac{s}{t}}[i-k]}[j-k]\right) + 1 = U_{\frac{s}{t}}[i-k, j-k] + 1.$$

□

Finally due to the symmetries in the second order balance matrix, we complete the construction of all the blocks of the matrix $U_{\frac{a}{b}}$.

Open questions After the detailed construction of the second order balance matrix, and the identification of each block, we can have some several questions.

1. First of all, what is the corresponding geometric signification of this matrix for the Christoffel words?

The answer of this question is given in Chapter 5.

2. Where can we use this order of balancedness and what information can we get from this matrix?

The first applications for this order of balancedness are given in Chapter 4 and open a wide door in research.

3. Moreover, as in Chapter 1, we were able to generalize the balance matrix and gives its form for a d -alphabet, we are still looking and searching for a general form for the matrix U_w , where w can be a binary word and not necessarily a 1-balanced word or even more, w can be a word over a d -alphabet.

Chapter 4

Some results

In Chapter 2, we talked about the isomorphism between the Christoffel tree and the Stern-Brocot tree that allows us to find the strong recursion in the composition of the matrix U_w . In this chapter, the work is divided into two parts, the first one shows some results about the second order balance $\delta^2(w)$. The second part explores the direct relation between the matrix U_w and the Christoffel word w without passing by the balance matrix B_w . In this part, we give some results concerning the maximal value of U_w or simply the second order balance. At the beginning, some calculations are made in order to deduce the algorithm of the increase of this second order in the Stern-Brocot tree. Then, with some computations, I was able to prove that this second order has a lower bound depending on its level on the Stern-Brocot tree and moreover, it can be determined depending on the values of the last two partial denominators of the slope a/b

4.1 Second order balance of $C(\frac{a}{b})$

After the construction of the second order balance matrix, and as we mentioned at the beginning of Chapter 3, we let the maximal value obtained in this matrix be the second order balance of $C(\frac{a}{b})$ which is denoted: $\delta^2(C(\frac{a}{b}))$. We let $\frac{a}{b} = [a_0, \dots, a_z]$, $\text{TBF}(\frac{a}{b}) = \frac{u}{v} = [a_0, \dots, a_{z-1} + 1]$, $\frac{c}{d} = [a_0, \dots, a_{z-1}, 2]$, $\frac{e}{f} = [a_0, \dots, a_{z-1} - 1, 2]$ and depending on the position of a/b in the Stern-Brocot tree, we let $\frac{\rho}{\theta} = \text{SUF}(\frac{a}{b})$.

We remark that this maximal value can be in one of these two blocks and is equal exactly to:

Theorem 4.1 *The second order balance of $C(\frac{a}{b})$ is:*

$$\delta^2\left(C\left(\frac{a}{b}\right)\right) = \max\left(\delta^2\left(C\left(\frac{u}{v}\right)\right), \delta^2\left(C\left(\frac{\rho}{\theta}\right)\right) + 1\right).$$

Proof. From Chapter 3, we knew that it is enough to detect the elements of the 3 blocks, α, β and γ to construct the matrix $U_{\frac{a}{b}}$ since the others are obtained by symmetry. The α -block is obtained from $U_{\frac{u}{v}}$ from lemma 3.10, the elements of the β -block are exactly the elements of $U_{\frac{e}{f}}$ increased by 1 from lemma 3.12 and finally, the γ -block is obtained from the γ -blocks of $u/v, c/d$ or e/f depending on the position of a/b on the Stern-Brocot tree from lemma 3.11. The maximal value of the matrix $U_{\frac{a}{b}}$ appears in one of these three blocks. We note that for both cases, $a_z = 2$ and $a_z \geq 3$, the maximal value of the γ -block of a/b is always less or equal to the maximal values of the α or β blocks. This is due to the fractions used to get this block and their positions on the Stern-Brocot tree. Therefore we get that $\delta^2(C(\frac{a}{b})) = \max(\delta^2(C(\frac{u}{v})), \delta^2(C(\frac{e}{f})) + 1)$, where the +1 is added since the elements of this block are the elements of $U_{\frac{e}{f}}$ increased by 1 from lemma 3.12. \square

We have mentioned before that the Stern-Brocot tree is divided into two symmetrical parts and we will be studying only the left part. The fractions on this tree can be located either on an extended branch or at the first deviation position. In other words, the last element of the continued fraction of each rational number can be strictly greater than 2 or equal to 2. We assign to each fraction on the first half of the tree its proper second order balance starting from the first fraction $\frac{1}{1}$ on the first level of the tree. Figure 4.1 shows the first 4 levels of the tree that contains the fractions and their second order balance value.

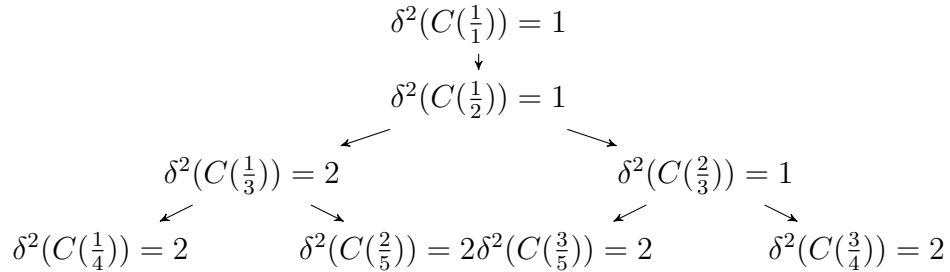


Figure 4.1: The first four levels of the Stern-Brocot tree containing the fractions and their second order balance value.

We can notice that depending on the position of the fraction on the tree, we are able to determine its second order balance value. Hence, the following lemmas study some particular cases in order to reach a general result that gives us a lower limit for the values of δ^2 on each level of the tree.

Lemma 4.2 *If the fraction a/b is a first deviation fraction with respect to u/v , then $\delta^2(C(\frac{a}{b})) \in \{\delta^2(C(\frac{u}{v})); \delta^2(C(\frac{u}{v})) + 1\}$.*

Proof. Let $a/b = \text{FEF}(u/v)$ where $\delta^2(C(\frac{u}{v})) = k$. From theorem 4.1, we have: $\delta^2(C(\frac{a}{b})) = \max(\delta^2(C(\frac{u}{v})), \delta^2(C(\frac{\rho}{\theta})) + 1)$ where $\frac{\rho}{\theta} = \text{SUF}(a/b)$. Depending on the position of the fraction $\frac{\rho}{\theta}$, $\delta^2(C(\frac{\rho}{\theta})) \in \{k, k - i\} \forall 1 \leq i \leq k - 1$. If $\delta^2(C(\frac{\rho}{\theta})) = k$ then $\delta^2(C(\frac{a}{b})) = k + 1$; Otherwise $k - i + 1 \leq k$ hence $\delta^2(C(\frac{a}{b})) = k$. \square

Lemma 4.3 *If the fraction $a/b = \text{FEF}(p/q)$, then $\delta^2(C(\frac{a}{b}))$ is increased each 2 steps.*

Proof. Let $a/b = \text{FEF}(p/q)$ with $\delta^2(C(\frac{p}{q})) = k$. We have 2 cases to study, either $\frac{\rho}{\theta} = \text{TBF}(\frac{a}{b}) = \frac{u}{v}$ where $\frac{\rho}{\theta} = \text{SUF}(a/b)$, or $\frac{\rho}{\theta} \neq \frac{u}{v}$.

- If $\frac{\rho}{\theta} = \frac{u}{v}$, then $\delta^2(C(\frac{a}{b})) = \max(\delta^2(C(\frac{u}{v})), \delta^2(C(\frac{\rho}{\theta})+1) = \max(\delta^2(C(\frac{\rho}{\theta})), \delta^2(C(\frac{\rho}{\theta})+1) = \delta^2(C(\frac{\rho}{\theta})) + 1$.
- If $\frac{\rho}{\theta} \neq \frac{u}{v}$, we have $\delta^2(C(\frac{u}{v})) < k$ hence, $\delta^2(C(\frac{a}{b})) = \max(\delta^2(C(\frac{u}{v})), \delta^2(C(\frac{\rho}{\theta})) + 1) = \delta^2(C(\frac{\rho}{\theta})) + 1$.

\square

We denote the second order balance of any fraction at level k in the Stern-Brocot tree by δ_k^2 and we have the following result:

Theorem 4.4 *For each level k on the Stern-Brocot tree, we have: $\delta_k^2 \geq \lceil \frac{k}{3} \rceil$.*

Proof. By induction on the levels of the Stern-Brocot tree, we have $\delta_1^2 = \delta_2^2 = 1$; $\delta_3^2 \in \{1, 2\}$ and $\delta_4^2 = 2$. Hence the minimal value of δ_3^2 ($\text{min}\delta_3^2$) is equal to 1 and $\text{min}\delta_4^2 = 2$. Therefore $\delta_i^2 \geq \lceil \frac{i}{3} \rceil$ with $1 \leq i \leq 4$.

Suppose that it is true for all the levels till the level $k = 3k'$; i.e $\text{min}\delta_k^2 = \text{min}\delta_{k-1}^2 = \text{min}\delta_{k-2}^2 = k'$. We prove that $\delta_{k+1}^2 \geq \lceil \frac{k+1}{3} \rceil = k' + 1$.

All the fractions on the level k with $\delta^2 > k'$ give fractions on the level $k + 1$ with $\delta^2 \geq k' + 1$. It remains to study the case of the fractions that have $\delta_k^2 = k'$ at level k . We let u/v be a fraction at level k with $\delta^2(C(\frac{u}{v})) = k'$. By contradiction, we let c/d be the fraction at level $k + 1$ with $\delta^2(C(\frac{c}{d})) = k'$. The fraction c/d is either a first deviation fraction with respect to u/v or is a first deviation position with respect to u/v . If c/d is the first extended fraction of u/v , then by lemma 4.3 we get $\delta^2(C(\frac{a}{b})) = k' - 1$, where a/b is the fraction at level $k - 1$. If c/d is the first deviation fraction with respect to u/v , then the fraction u/v is either a first extended fraction of a/b or a first deviation fraction with respect to a/b where m/n is the fraction at level $(k - 2)$.

- If $u/v = \text{FEF}(a/b)$ then by lemma 4.3, we get $\delta^2(C(\frac{m}{n})) = k' - 1$.
- If $u/v = \text{FDF}(a/b)$, then if $\delta^2(C(\frac{a}{b})) = k'$, we get:
 $\delta^2(C(\frac{c}{d})) = \max(\delta^2(C(\frac{u}{v})), \delta^2(C(\frac{m}{n})) + 1) = k'$. If $\delta^2(C(\frac{m}{n})) = k'$ then
 $\delta^2(C(\frac{c}{d})) = k' + 1$; and if $\delta^2(C(\frac{m}{n})) = k' - 1$ we get a contradiction since
 $\delta_{(k-2)}^2 \geq \lceil \frac{k-2}{3} \rceil = k'$.

Therefore $\delta^2 \geq k' + 1$ for all the fractions on the level $k + 1$ thus

$$\delta_{k+1}^2 \geq \lceil \frac{k+1}{3} \rceil = \lceil \frac{3k'+1}{3} \rceil = k' + 1.$$

□

We have several paths on the Stern-Brocot tree, a particular path called the *Zigzag* path is introduced in the following definition. This path has a particular property that will be discussed in the next paragraph.

Definition 4.5 *We define the zigzag path to be the path followed by the fractions in the Stern-Brocot tree having $[1, 1, 1, \dots]$ as continued fraction.*

The following proposition gives the recursion used to get the values of the second order balance of all the fractions on the Zigzag path. We notice that the inequality in Theorem 4.4 is also achieved.

Proposition 4.6 *In the zigzag path, the value of δ^2 is increased by 1 after each 3 levels.*

Proof. By induction on the zigzag path, we have $\delta^2(C(\frac{1}{1})) = \delta^2(C(\frac{1}{2})) = \delta^2(C(\frac{2}{3})) = 1$, while $\delta^2(C(\frac{3}{5})) = 2$. We suppose that it is true till level k and we prove it true for level $k + 1$. Hence we let $a/b, c/d, e/f$ be 3 consecutive fractions on the zigzag path such that: $\delta^2(C(\frac{a}{b})) = \delta^2(C(\frac{c}{d})) = \delta^2(C(\frac{e}{f})) = k'$. We prove that the following fraction g/h on this path has: $\delta^2(C(\frac{g}{h})) = k' + 1$. We have $\text{TBF}(\frac{g}{h}) = \frac{e}{f}$ and the fraction $a/b = \text{SUF}(g/h)$, therefore: $\delta^2(C(\frac{g}{h})) = \max(\delta^2(C(\frac{e}{f})), \delta^2(C(\frac{a}{b})) + 1) = k' + 1$. □

4.1.1 Refinement of the value of $\delta^2(\frac{a}{b})$ using the continued fraction

We noticed that the value of δ^2 increases every two or three levels in the Stern-Brocot tree depending on the path followed to reach the rational number. As

presented in Chapter 2, we can acquire the continued fraction of each rational number. Hence, there is certainly a relation between the value attained by δ^2 and the partial denominators of the rational number a/b .

From Theorem 4.1, we know that $\delta^2 \left(C \left(\frac{a}{b} \right) \right) = \max \left(\delta^2 \left(C \left(\frac{u}{v} \right) \right), \delta^2 \left(C \left(\frac{\rho}{\theta} \right) \right) + 1 \right)$, where $\frac{u}{v} = \text{TBF}(\frac{a}{b})$; $\frac{\rho}{\theta} = \text{SUF}(\frac{a}{b})$. From Definition 2.28, we recall:

- $\frac{a}{b} = [a_0, a_1, \dots, a_z]$;
- $\frac{u}{v} = [a_0, a_1, \dots, a_{z-1}, 1] = [a_0, a_1, \dots, a_{z-1} + 1]$;
- $\frac{\rho}{q} = [a_0, \dots, a_{z-2}, 1] = [a_0, \dots, a_{z-2} + 1]$;
- $\frac{s}{t} = [a_0, \dots, a_{z-3}]$;
- $\frac{\rho}{\theta} = \begin{cases} [a_0, a_1, \dots, a_z - 2] & \text{if } a_z \geq 4 \\ [a_0, a_1, \dots, a_{z-1} + 1] & \text{if } a_z = 3 \\ [a_0, a_1, \dots, a_{z-2}] & \text{if } a_z = 2. \end{cases}$

Without loss of generality and in order to enlighten the notation, we denote $\delta^2 \left(C \left(\frac{a}{b} \right) \right)$ by D_a , $\delta^2 \left(C \left(\frac{u}{v} \right) \right) = D_u$, $\delta^2 \left(C \left(\frac{\rho}{\theta} \right) \right) = D_r$, $\delta^2 \left(C \left(\frac{\rho}{q} \right) \right) = D_p$ and $\delta^2 \left(C \left(\frac{s}{t} \right) \right) = D_s$.

In the following theorem, we discuss all the possible cases for the distribution of the fractions on the Stern-Brocot tree. We deduce that the value of $\delta^2 \left(C \left(\frac{a}{b} \right) \right)$ is always equal to the value of $\delta^2 \left(C \left(\frac{\rho}{\theta} \right) \right) + 1$, except for the case when $a_{z-1} \geq 2$.

Theorem 4.7 For $\frac{a}{b} = [a_0, a_1, \dots, a_z]$, we have:

$$\delta^2 \left(C \left(\frac{a}{b} \right) \right) = \begin{cases} D_u = \delta^2 \left(C \left(\frac{u}{v} \right) \right) & \text{if } a_{z-1} \geq 2 \\ D_r + 1 = \delta^2 \left(C \left(\frac{\rho}{\theta} \right) \right) + 1 & \text{Elsewhere.} \end{cases}$$

Proof. The proof of this theorem is divided into 3 parts that discuss all the possible paths to reach the fraction a/b . Since $\frac{a}{b} = [a_0, a_1, \dots, a_z]$, we build our study and discussion on the different possible values of a_z , a_{z-1} and a_{z-2} , and we get:

1. If $a_z \geq 4$, we let $D_u = d$, then $D_r \geq d$ since $\frac{\rho}{\theta} > \frac{u}{v}$. Hence $D_a = \max(D_u, D_r + 1) = D_r + 1$. (see Figure 4.2).
2. If $a_z = 3$, in this case the fractions $\frac{u}{v}$ and $\frac{\rho}{\theta}$ are the same and we have: $D_a = \max(D_u, D_r + 1) = \max(D_r, D_r + 1) = D_r + 1$. (see Figure 4.2).

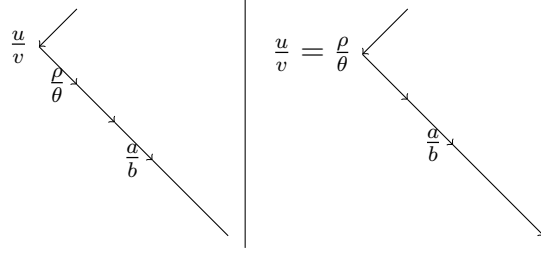


Figure 4.2: Case where $a_z \geq 4$ and $a_z = 3$ respectively.

3. If $a_z = 2$, in this case we get several sub-cases depending on the value of a_{z-1} and a_{z-2} . For all the cases, we have: $\frac{u}{v} > \frac{\rho}{\theta}$. In this case, we recall: $\frac{a}{b} = [a_0, \dots, 2]$; $\frac{u}{v} = [a_0, \dots, a_{z-1} + 1]$ and $\frac{\rho}{\theta} = [a_0, \dots, a_{z-2}]$.

(a) If $a_{z-1} = 1$, we get the fraction $\frac{u}{v} = [a_0, \dots, 2]$ and we let $D_r = d$. Figure 4.3 shows the positions of the fractions in this case along the path. We can note that 4 sub-cases arise and they are summarized in Table 4.1.

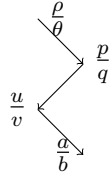


Figure 4.3: Geometric representation of the path having $a_z = 2$ and $a_{z-1} = 1$.

D_r	d			
D_p	d	$d + 1$		
D_u	d	$d + 1$	$d + 1$	$d + 2$
D_a	$d + 1$	$d + 1$	$d + 1$	X

Table 4.1: The representation of the 4 cases for $a_{z-1} = 1$.

- i- If $D_p = d$, then $D_u = d$ or $D_u = d + 1$.
 If $D_u = d$, then $D_a = \max(D_u, D_r + 1) = \max(d, d + 1) = d + 1 = D_r + 1$,
 If $D_u = d + 1$, then $D_a = \max(D_u, D_r + 1) = \max(d + 1, d + 1) = d + 1 = D_r + 1$.
- ii- If $D_p = d + 1$, then $D_u = d + 1$ or $D_u = d + 2$.
 If $D_u = d + 1$, then $D_a = d + 1 = D_r + 1$, already studied,
 If $D_u = d + 2$, we get a contradiction because in this case:

$$D_u = \max(D_p, D_s + 1)$$

$$d + 2 = \max(d + 1, D_s + 1),$$

hence $D_s = d + 1$ which it can not be since $d_r = d$ and we have $\frac{s}{t} < \frac{p}{q}$.

- (b) If $a_{z-1} = 2$, the fractions are distributed in two possible ways depending on the value of a_{z-2} that can be equal to 2 or greater or equal to 3. The positions of the fractions are represented in Figure 4.2. We remark that the result is obtained depending on the value of D_p . The cases are represented in Table 4.4 that is true for any value of $a_{z-2} \geq 1$.

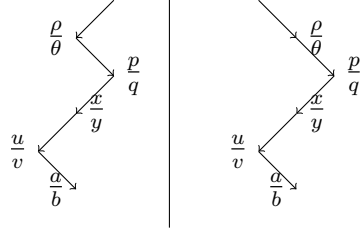


Figure 4.4: Geometric representation of the path having $a_z = 2$ and $a_{z-1} = 2$.

- If $D_p = d$, then by Lemma 4.3, $D_u = d + 1$ and $D_a = \max(D_u, D_r + 1) = \max(d + 1, d + 1) = d + 1 = D_r + 1$,
- If $D_p = d + 1$, then $D_u = d + 2$ and $D_a = \max(D_u, D_r + 1) = \max(d + 2, d + 1) = d + 2 = D_u$.

D_r	d	
D_p	d	$d + 1$
D_u	$d + 1$	$d + 2$
D_a	$d + 1$	$d + 2$

Table 4.2: The representation of the 2 cases for $a_{z-1} = 2$.

From this calculation, we remark that we always have $D_a = D_u$ regardless the value of a_{z-2} .

- (c) Last case remained to study is when $a_{z-1} \geq 3$,
In fact, if $a_{z-1} > 3$, we always have $D_u > D_r + 1$ therefore, $D_a = D_u$. See Figure 4.5.

While for the case where $a_{z-1} = 3$, Figure 4.6, represents the geometric form of the path followed to reach the rational number a/b having $a_{z-1} = 3$. We let $\frac{k}{l} = \text{FEF}\left(\frac{p}{q}\right)$ with $\delta^2\left(C\left(\frac{k}{l}\right)\right) = D_k$. In fact the

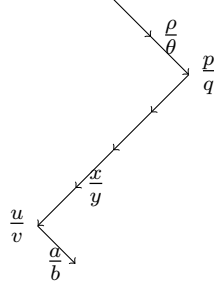


Figure 4.5: Geometric representation for the case $a_{z-1} = 4$.

value of D_a depends on the value of D_k and we have two cases that are represented in Table 4.3.

- i If $D_k = d$, then by Lemma 4.3, $D_u = d + 1$ and $D_a = \max(D_u, D_r + 1) = \max(d + 1, d + 1) = d + 1 = D_r + 1$.
- ii If $D_k = d + 1$, then by Lemma 4.3, $D_u = d + 2$ and $D_a = \max(D_u, D_r + 1) = \max(d + 2, d + 1) = d + 2 = D_u$.

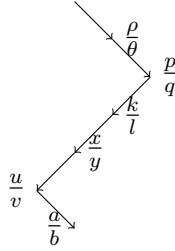


Figure 4.6: Geometric representation for the path $a_{z-1} = 3$.

D_r	d	
D_k	d	$d + 1$
D_u	$d + 1$	$d + 2$
D_a	$d + 1$	$d + 2$

Table 4.3: The representation of the 2 cases for $a_{z-1} = 3$.

Similarly in this case, we also have that $D_a = D_u$.

□

4.1.2 Fibonacci sequence

In the 18th century, the Italian mathematician "Leonardo di Pisa" has introduced the following substitution on the binary alphabet $A = \{0, 1\}$ that gave what is called now the Fibonacci substitution:

$$\sigma : A \rightarrow A^* \begin{cases} 0 \rightarrow 01 \\ 1 \rightarrow 0 \end{cases}$$

It follows that $\sigma(w_1w_2 \dots w_n) = \sigma(w_1)\sigma(w_2) \dots \sigma(w_n)$.

Example 65 Let $w = 00100101 \in A^*$, using the Fibonacci substitution we get:

$$\sigma(w) = \sigma(0)\sigma(0)\sigma(1)\sigma(0)\sigma(0)\sigma(1)\sigma(0)\sigma(1) = 0101001010010.$$

By applying several iterations on $\sigma(0)$ i.e by calculating each time $\sigma^i(0) = \sigma(\sigma^{i-1}(0))$ we get the Fibonacci word which is obtained when $\sigma(X) = X$.

Example 66 Let us calculate the first 6 iterations on $\sigma(0)$, we get:

$$\begin{aligned} \sigma(0) &= 01 \\ \sigma^2(0) &= \sigma(01) = 010 \\ \sigma^3(0) &= \sigma(\sigma^2(0)) = \sigma(010) = 01001 \\ \sigma^4(0) &= \sigma(\sigma^3(0)) = \sigma(01001) = 01001010 \\ \sigma^5(0) &= \sigma(\sigma^4(0)) = \sigma(01001010) = 0100101001001 \\ \sigma^6(0) &= \sigma(\sigma^5(0)) = \sigma(0100101001001) = 010010100100101001010. \end{aligned}$$

For each step we calculate the length of the word obtained, we suppose $|\sigma^0(0)| = |\sigma^{-1}(0)| = 1$ and we obtain the following sequence: 1 1 2 3 5 8 ... that respects the FIBONACCI relation:

$$|\sigma^{n+2}(0)| = |\sigma^{n+1}(0)| + |\sigma^n(0)| \quad \forall n \geq 0.$$

Example 67 From Example 66, we get: $|\sigma(0)| = 2, |\sigma^2(0)| = 3, |\sigma^3(0)| = 5, |\sigma^4(0)| = 8, |\sigma^5(0)| = 13$ and $|\sigma^6(0)| = 21$. We remark that the Fibonacci relation is respected since:

$$\begin{aligned} 5 &= |\sigma^3(0)| = |\sigma^2(0)| + |\sigma(0)| = 3 + 2 \\ 8 &= |\sigma^4(0)| = |\sigma^3(0)| + |\sigma^2(0)| = 5 + 3 \\ 13 &= |\sigma^5(0)| = |\sigma^4(0)| + |\sigma^3(0)| = 8 + 5 \\ 21 &= |\sigma^6(0)| = |\sigma^5(0)| + |\sigma^4(0)| = 13 + 8. \end{aligned}$$

Two points are to be noticed:

1. We can note that $\frac{a_n}{b_n} = \frac{|\sigma^{n+1}(0)|_1}{|\sigma^{n+1}(0)|_0}$.
2. By taking $\frac{a_n}{b_n} = \frac{|\sigma^{n-1}(0)|}{|\sigma^n(0)|}$, we get the zigzag path in the Stern-Brocot tree defined in Definition 4.5.

Example 68 For the word $w' = \sigma^6(0) = 010010100100101001010$, we have $|w'|_1 = 8$ and $|w'|_0 = 13$, therefore we can write: $\frac{8}{13} = \frac{|w'|_1}{|w'|_0} = \frac{|\sigma^4(0)|}{|\sigma^5(0)|}$.

By proposition 4.6 we see that the zigzag path associated with $[1, 1, 1 \dots]$ minimizes the growth of δ^2 in the stern-Brocot tree. We have:

$$\lim_{n \rightarrow +\infty} \frac{a_n}{b_n} = \varphi = \frac{1 + \sqrt{5}}{2}.$$

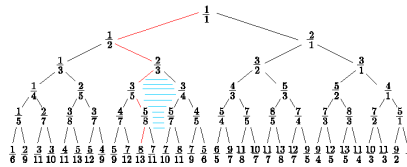


Figure 4.7: The Fibonacci sequence in the Stern-Brocot tree.

Open questions The zigzag path is one of the paths that minimize the growth of δ^2 in the Stern-Brocot tree, is it the only path?

The following code allows us to answer this question negatively but pushes us to ask more questions around this δ^2 .

1. What are the paths that minimize the growth of δ^2 all over the Stern-Brocot tree?
2. Are there any particular paths in this tree like the Fibonacci path?
3. What are the characteristics or particularities for all those paths?
4. By calculating the partitions of a fixed integer n belonging to the Fibonacci sequence, the zigzag path is one of the paths that minimizes the growth of δ^2 . Therefore the following question arises: By calculating the partition of a fixed value $n = a + b$, where n is not an element of the Fibonacci sequence, what are the paths that minimize the value of δ^2 ?

All those questions can be asked and still points and ideas under studies. Code 4.1 gives us all the minimal paths on the Stern-Brocot tree with all the fractions along the path. While Code 4.2 gives us the value of δ^2 for all the fractions of same length.

```

1 def S(n):
2     if n==0:
3         return 1
4     if n==1:
5         return 2
6     l=n.bit_length()-1
7     r=n-2**l
8     if r>= 0 and r < 2**(l-1):
9         return S(2**(l-1)+r)+S(r)
10    else:
11        return S(2**(l+1)-r-1)
12 def val_2(L):
13     s=""
14     for i in L:
15         s+=str(i)
16     return int(s,2)
17 def fraction(L):
18     L_1=L[1:]
19     a=S(val_2(L_1))
20     b=S(val_2(L))
21     return ((a,b))
22 def cheminx_min(k):
23     F=[[ (1,1) ,(1,2) ]]
24     W=[[ 1,1]]
25     WW=[]
26     for i in range (k):
27         FF=[]
28         for w in W:
29             w_1=w+[0]
30             w_2=w+[1]
31             f_1=fraction(w_1)
32             f_2=fraction(w_2)
33             d_1=secondOrderBalancePython(f_1[0],f_1[1])
34             d_2=secondOrderBalancePython(f_2[0],f_2[1])
35             if d_1==ceil((i+1)/3):
36                 WW.append(w_1)
37                 FF.append(f_1)
38 def exemple_iterator(a):
39     yield 1
40     yield 2
41     yield 3
42 def word_generator(k):
43     if k == 0:

```

```

44     yield tuple()
45     else:
46         for w in word_generator(k-1):
47             yield w + (0,)
48             yield w + (1,)
49 def word_generator2(k):
50     if k == 0:
51         yield []
52     else:
53         for w in word_generator2(k-1):
54             yield w + [0]
55             yield w + [1]
56 def check_minimal(word, minimal_delta2):
57     f = fraction(word)
58     d = secondOrderBalancePython (f[0],f[1])
59     return d == minimal_delta2
60 def minimum_path(k):
61     if k < 2:
62         return
63     if k == 2:
64         yield [1,1]
65     else:
66         minimal_delta2 = ceil((k+1)/3)
67         for w in minimum_path(k-1):
68             w0 = w + [0]
69             w1 = w + [1]
70             if check_minimal(w0,minimal_delta2):
71                 yield w0
72             if check_minimal(w1,minimal_delta2):
73                 yield w1
74
75 def all_fractions(min_paths):
76     return [ [fraction(w[:i]) for i in xrange(2,len(w)+1)] for w in
min_paths]

```

Listing 4.1: Code of the paths that minimize the growth of δ^2 on the Stern-Brocot tree.

```

1 def fractionsQuiMinimisentSOB(n):
2     lesFractionsMin = []
3     valeurMinDelta2 = oo
4     F=[ (i,n-i) for i in range( 1, (n+1)//2 ) if pgcd( i, n-i ) == 1
5     ]
6     for a,b in F :
7         delta2 = max( max(i) for i in f(a,b) ) )
8         if delta2 < valeurMinDelta2 :
9             valeurMinDelta2 = delta2
10            lesfractionsMin = [ (a,b) ]

```

```

10     elif delta2 == valeurMinDelta2 :
11         lesFractionsMin.append( (a,b) )
12     return valeurMinDelta2 , lesFractionsMin

```

Listing 4.2: Minimal value of δ^2 for a fixed length.

4.2 Abelian vectors

In the main definition used to compute the second order balance matrix, we pass through the balance matrix first and then test the balancedness of each row of this matrix. In fact a normal question derives, aren't we able to obtain the entries of the second order balance matrix immediately from the Christoffel word $p = C(\frac{a}{b})$ without using the rows of B_p ? This was an interesting question that pushes us to look for the direct relation between U_p and p . The relation is found but before introducing it, we need to prove in Lemma 4.8 that for any factor w of p of length $i + j - 1$ where $1 \leq i, j \leq n - 1$, the number of occurrences of 1 is conserved if we decompose w into j consecutive factors of length i or into i consecutive factors of length j .

We denote V_i^j the list composed of j consecutive factors of length i each, and U_j^i the set of i consecutive factors of length j each.

Example 69 Let us consider the binary word $w = 1001001$ of length 7. We compute the two lists: V_3^5 and U_5^3 where $i = 3, j = 5$ and $i + j - 1 = 7 = |w|$.

- $V_3^5 = (100, 001, 010, 100, 001)$,
- $U_5^3 = (10010, 00100, 01001)$.

Lemma 4.8 Let w be a factor of the Christoffel word $p = C(\frac{a}{b})$ such that $|w| = i + j - 1$ and $w[1 \dots i + j - 1] = p[k \dots k + i + j - 2]$. If the $V_i^j = (v_i^l = w[l \dots l + i - 1])_{l=1 \dots j}$ and $U_j^i = (u_j^l = w[l \dots l + j - 1])_{l=1 \dots i}$, then:

$$\sum_{l=1}^j |v_i^l|_1 = \sum_{l=1}^i |u_j^l|_1.$$

Proof. Without loss of generality, we let $j < i$ and we write $tj = i + \gamma$ where $0 \leq \gamma < j$ and $t \geq 1$. We get the following factors:

V_i^j	U_j^i
$v_i^1 = w[1 \dots i]$	$u_j^1 = w[1 \dots j]$
$v_i^2 = w[2 \dots i + 1]$	$u_j^2 = w[2 \dots j + 1]$
\vdots	\vdots
$v_i^j = w[j \dots i + j - 1]$	$u_j^i = w[i \dots i + j - 1]$

We decompose all the factors of length i into small factors of length j and we see that all the factors of length j are used to cover the factors of length i .

$v_i^1 =$	$w[1 \dots \dots \dots i]$	
$=$	$w[1 \dots j]$	$w[(t-1)j+1 \dots tj]$
$=$	$w[1 \dots j]$	$w[(t-1)j+1 \dots tj - \gamma] w[i+1 \dots tj]$

Table 4.4: Decomposition of an element of V_i^j by factors of length j .

We remark that the element v_i^1 is covered by some elements of U_j^i . In particular, we use the elements $u_j^1; u_j^{j+1} \dots u_j^{(t-1)j+1}$. By concatenating these factors the element $u_j^{(t-1)j+1}$ exceeded the length of v_i^1 with a factor of length γ . Hence we decompose it into two parts and we name the remaining part $A_1 = w[i+1 \dots tj]$. Therefore, we have:

$$|v_i^1|_1 = \sum_{l=1}^t |u_j^{(l-1)j+1}|_1 - |A_1|_1.$$

Similarly for the factor v_i^2 , we can decompose it in the same way using exactly t vectors of U_j^i and having the factor $A_2 = w[i+2 \dots tj+1]$ of length γ as an exceeded factor. Therefore

$$|v_i^2|_1 = \sum_{l=1}^t |u_j^{(l-1)j+2}|_1 - |A_2|_1.$$

Same reasoning is followed till the factor $v_i^{j-\gamma}$ and we have:

$$A_s = w[i+s \dots tj+s-1], \forall 1 \leq s \leq j-\gamma.$$

While for the element $v_i^{j-\gamma+1}$, we can write:

$w[j-\gamma+1 \dots \dots \dots i+j-\gamma]$		
$w[j-\gamma+1 \dots 2j-\gamma]$	$w[(t-2)j+j-\gamma+1 \dots (t-1)j+j-\gamma]$	$w[i+1 \dots i+j-\gamma]$
$w[j-\gamma+1 \dots 2j-\gamma]$	$w[i-j \dots \dots \dots i]$	$w[i+1 \dots i+j-\gamma]$

Table 4.5: Decomposition of the element $v_i^{j-\gamma+1}$ by factors of length j .

In fact we have: $(t-1)j+j-\gamma = i$ and since $v_i^{j-\gamma}$ is not fully covered we still need to start by a new factor starting from $i+1$. Since we don't have any factor

of U_j^i that starts with this value, therefore

$$|v_i^{j-\gamma+1}|_1 = \sum_{l=1}^{t-1} |u_j^{(l-1)j+j-\gamma}|_1 + |B_1|_1,$$

where $B_1 = w[i+1 \dots i+j-\gamma]$ and $|B_1| = j-\gamma$. We repeat the same decomposition until we reach v_i^j , where we get:

$$|v_i^j|_1 = \sum_{l=1}^{t-1} |u_j^l|_1 + |B_\gamma|_1,$$

where $B_\gamma = w[i+\gamma-1 \dots i+j-1]$. Hence, we can write:

$$B_{s'} = w[i+s' \dots i+j-\gamma+s'-1], \forall 1 \leq s' \leq \gamma.$$

If $j = 2\gamma$, we get $j-\gamma = \gamma$ hence $A_s = B_{s'}$; $\forall 1 \leq s, s' \leq \gamma$. Otherwise, we remark that the factors A_s cover all the factors of $B_{s'}$. Hence we have:

$$\sum_{s=1}^{j-\gamma} |A_s|_1 = \sum_{s'=1}^{\gamma} |B_{s'}|_1.$$

By adding the two members we get:

$$\begin{array}{rcl} |v_i^1|_1 & = & \sum_{l=1}^t |u_j^{(l-1)j+1}|_1 - |A_1|_1 \\ |v_i^2|_1 & = & \sum_{l=1}^t |u_j^{(l-1)j+2}|_1 - |A_2|_1 \\ \vdots & & \vdots \\ |v_i^{j-\gamma+1}|_1 & = & \sum_{l=1}^{t-1} |u_j^{(l-1)j+j-\gamma}|_1 + |B_1|_1 \\ \vdots & & \vdots \\ |v_i^j|_1 & = & \sum_{l=1}^{t-1} |u_j^l|_1 + |B_\gamma|_1 \\ \hline \sum_{l=1}^j |v_i^l|_1 & = & \sum_{l=1}^t |u_j^l|_1. \end{array}$$

□

Example 70 Let $p = C(\frac{7}{18}) = 0001000100100010010001001$ and $w = p[7, 19] = 010010001001$, where $|w| = 13 = 9 + 5 - 1$. We let $i = 9$ and $j = 5$ and we have:

$$\begin{aligned}
V_9^5 &= (w[1, 9], w[2, 10], w[3, 11], w[4, 12], w[5, 13]) \\
&= (p[7, 15], p[8, 16], p[9, 17], p[10, 18], p[11, 19]) \\
&= (010010001, 100100010, 001000100, 010001001, 100010010) \\
|V_9^5|_1 &= (3, 3, 2, 3, 3) \\
U_5^9 &= (w[1, 5], w[2, 6], w[3, 7], w[4, 8], w[5, 9], w[6, 10], w[7, 11], w[8, 12], w[9, 13]) \\
&= (p[7, 11], p[8, 12], p[9, 13], p[10, 14], p[11, 15], p[12, 16], p[13, 17], p[14, 18], p[15, 19]) \\
&= (01001, 10010, 00100, 01000, 10001, 00010, 00100, 01001, 10010) \\
|U_5^9|_1 &= (2, 2, 1, 1, 2, 1, 1, 2, 2).
\end{aligned}$$

Where: $3 + 3 + 2 + 3 + 3 = 2 + 2 + 1 + 1 + 2 + 1 + 1 + 2 + 2 = 14$.

In this part we denote by Ψ , the morphism from $A^* \rightarrow \mathbb{Z}^2$ defined by: $\Psi(w) = (|w|_0, |w|_1)$. This vector is called the Parikh vector.

Definition 4.9 Let $p = C(\frac{a}{b})$ with $n = a + b$, $v_1, v_2 \in L_{i+j-1}(p)$ where $1 \leq i, j \leq n - 1$. We let $v_1^* = (v_1[l \cdots l + j - 1])_{(l=1 \cdots i)}$; $v_2^* = (v_2[l \cdots l + j - 1])_{(l=1 \cdots i)}$ and $v \in L_j(p)$. The Parikh vector of v_1^* and v_2^* are: $\Psi(v_1^*) = (\Psi(v_1[l \cdots l + j - 1]))_{(l=1 \cdots i)}$ and $\Psi(v_2^*) = (\Psi(v_2[l \cdots l + j - 1]))_{(l=1 \cdots i)}$. We define the ABELIAN matrix as follows:

$$AB_{\frac{a}{b}}[i, j] = \max_{\substack{v_1 \in L_{i+j-1}(p) \\ v \in L_j(p)}} |\Psi(v_1^*)|_{\Psi(v)} - \min_{\substack{v_2 \in L_{i+j-1}(p) \\ v \in L_j(p)}} |\Psi(v_2^*)|_{\Psi(v)},$$

where $|\Psi(v_1^*)|_{\Psi(v)}$ is the number of occurrences of the couple $\Psi(v)$, which is $(|w|_0, |w|_1)$, in the list of couples $\Psi(v_1^*)$.

In another words, we are counting the maximal difference of the occurrences of the abelian vector of v in the list of all the abelian factors of v_1 of length j where v_1 is of length $i + j - 1$; $\forall 1 \leq i, j \leq n - 1$.

In Theorem 4.10, we prove that the Abelian matrix and the second order balance matrix for the Christoffel word of slope a/b are identical.

Theorem 4.10 Let $p = C(\frac{a}{b})$ be a Christoffel word of slope a/b with $n = a + b$. The second order balance matrix of a/b is equal to the Abelian matrix of a/b , i.e

$$U_{\frac{a}{b}}[i, j] = AB_{\frac{a}{b}}[i, j], \quad \forall 1 \leq i, j \leq n - 1.$$

Proof. From the previous chapters, we know that the second order balance matrix is given by:

$$U_{\frac{a}{b}}[i, j] = \max_{v_1 \in L_j(B_{\frac{a}{b}}[i])} |v_1|_1 - \min_{v_2 \in L_j(B_{\frac{a}{b}}[i])} |v_2|_1$$

with $B_{\frac{a}{b}}[i, j] = S_{\frac{a}{b}}[i, j] - \min(S_{\frac{a}{b}}[i])$ where $S_{\frac{a}{b}}[i, j] = |p[j \dots i + j - 1]|_1$.

Let v_1 and v_2 be two vectors in $L_j(B_{\frac{a}{b}}[i])$, where $|v_1| = |v_2| = j$, such that v_1 starts at position k and v_2 at position l . We suppose that v_1 is the factor of the maximal value and v_2 is the factor of the minimal value. We consider the circular Christoffel word p^ω , and since v_1, v_2 are factors of $B_{\frac{a}{b}}[i]$ then we can decompose the factors as follow:

$$\begin{aligned} |v_1[1]|_1 &= |p[k \dots k + i - 1]|_1 - \min(S_{\frac{a}{b}}[i]) \\ |v_1[2]|_1 &= |p[k + 1 \dots k + i]|_1 - \min(S_{\frac{a}{b}}[i]) \\ &\vdots \\ |v_1[j]|_1 &= |p[k + j - 1 \dots k + i + j - 2]|_1 - \min(S_{\frac{a}{b}}[i]) \end{aligned}$$

$$\begin{aligned} |v_2[1]|_1 &= |p[l \dots l + i - 1]|_1 - \min(S_{\frac{a}{b}}[i]) \\ |v_2[2]|_1 &= |p[l + 1 \dots l + i]|_1 - \min(S_{\frac{a}{b}}[i]) \\ &\vdots \\ |v_2[j]|_1 &= |p[l + j - 1 \dots l + i + j - 2]|_1 - \min(S_{\frac{a}{b}}[i]) \end{aligned}$$

Since v_1 and v_2 are factors of $B_{\frac{a}{b}}[i]$, then we have the same $\min(S_{\frac{a}{b}}[i])$ which allows us to write:

$$\begin{aligned} U_{\frac{a}{b}}[i, j] &= \sum_{t=1}^j |v_1[t]|_1 - \sum_{t=1}^j |v_2[t]|_1 \\ &= \sum_{t=1}^j |p[k + t - 1 \dots k + i + t - 2]|_1 - \sum_{t=1}^j |p[l + t - 1 \dots l + i + t - 2]|_1 \\ &= z. \end{aligned}$$

In other words, $z = \sum_{t=1}^j |w_i^t|_1 - \sum_{t=1}^j |w_i''^t|_1$, where w_i^t are the j consecutive factors of length i of the word $w' = p[k \dots k + i + j - 2]$ and $w_i''^t$ are for the factor

$w'' = p[l \dots l + i + j - 2]$, with $|w'| = |w''| = i + j - 1$.

In another hand, we let the two factors $w', w'' \in L_{i+j-1}(p)$ such that $w' = p[k \dots k + i + j - 2]$ and $w'' = p[l \dots l + i + j - 2]$. We let $(w')^* = (w'[l \dots l + j - 1])_{(l=1 \dots i)}$, the vector of all the factors of w' and of length j . Same process is repeated with the factor w'' . We define the vector A_1 to be the vector corresponding to the $\Psi((w')^*)$ which is the vector that gives the Parikh vector for each factor in $(w')^*$. Similarly, we let A_2 be the Parikh vector of $(w'')^*$. We know that Christoffel words are standard strumian words with abelian complexity equal to 2 by Zamboni [44]. Therefore we can find only two distinct elements in A_1 and A_2 , where the cardinality of each is equal to i . Therefore, we let $(r_1, r_2); (n_1, n_2)$ be the two distinct elements of A_1 and A_2 . We let k_A (resp. l_A) be the number of (r_1, r_2) (resp. (n_1, n_2)) in A_1 and k_{A_2} (resp. l_{A_2}) be the number of (r_1, r_2) (resp. (n_1, n_2)) in A_2 . We have the following relations:

- $n_1 + n_2 = r_1 + r_2 = j$.
- $k_A + l_A = k_{A_2} + l_{A_2} = i$.
- Since p is a 1-balanced Christoffel word then we have: $r_2 = n_2 + 1$ and $r_1 = n_1 - 1$.

Therefore, if we let $k_A - k_{A_2} = l_A - l_{A_2} = u$, and by using the lemma 4.8, we get:

$$\begin{aligned}
 z &= k_A r_2 + l_A n_2 - k_{A_2} r_2 - l_{A_2} n_2 \\
 &= (k_A - k_{A_2}) r_2 + (l_A - l_{A_2}) n_2 \\
 &= u r_2 - u n_2 \\
 &= u(r_2 - n_2) \\
 &= u.
 \end{aligned}$$

Therefore,

$$U_{\frac{a}{b}}[i, j] = AB_{\frac{a}{b}}[i, j].$$

□

Sage code The code 4.3 is used to determine the matrix AB .

```

1 import numpy as np
2 def iter (w):
3     w=Word(w, datatype="str")
4     W=Words('01')
5     IT=[]

```

```

6     l2=[]
7     n=len(w)
8     for j in range(1,n+1):
9         SPP=[]
10        factt=[]
11        for i in range(n-j+1):
12            SPP.append(w[i:i+j])
13            factt.append(W(w[i:i+j]).abelian_vector())
14        IT.append(SPP)
15        l2.append(factt)
16
17    return(IT,l2)
18 def AB(a,b):
19     cri=Word(words.ChristoffelWord(a,b), datatype="list")
20     cri=cri+cri
21     #cri=Word(cri)
22     n=a+b
23     L=[]
24     W=Words('01')
25     for k in range(n+n+1): L.append(cri.factor_set(k));
26     ABL=[]
27     for i in range(1,n+n+1):
28         ligne=[]
29         for j in range(1,n+1):
30             if j <= i:
31                 XJ=[]
32                 for w in L[i]:
33                     FF=[]
34                     for k in range(j+1):
35                         FF.append(iter(w)[1][j-1].count([k,j-k]))
36                     XJ.append(FF)
37                 fin=[]
38                 for t in range(j+1):
39                     suf=[]
40                     for s in range(len(L[i])):
41                         suf.append(XJ[s][t])
42                     m_1=max(suf)
43                     m_2=min(suf)
44                     fin.append(m_1-m_2)
45                 sol=max(fin)
46                 ligne.append(sol)
47             else:
48                 ligne.append(0)
49         ABL.append(ligne)
50     B=[[0 for i in range(n-1)] for j in range(n-1)]
51     for i in range(n-1):
52     for j in range(n-1):
53         B[i][j]=ABL[i+j][j]

```

```

54 B=np.array(B)
55 return B

```

Listing 4.3: Code of the matrix AB .

Example 71 Let $w = C(\frac{4}{9})$, Table 4.6 represents the Parikh vectors of length 5 for all the factors of length 9 of w . Therefore $AB_w[5, 5] = \max(R_1) = 3$.

	(5, 0)	(4, 1)	(3, 2)	(2, 3)	(1, 4)	(0, 5)
001000100	0	4	1	0	0	0
001001001	0	2	3	0	0	0
001001000	0	3	2	0	0	0
010001001	0	3	2	0	0	0
000100100	0	3	2	0	0	0
010010001	0	2	3	0	0	0
010010010	0	1	4	0	0	0
100010010	0	2	3	0	0	0
100100010	0	3	2	0	0	0
100100100	0	2	3	0	0	0
$R_1 = \max - \min$	0	3	3	0	0	0

Table 4.6: Table of Parikh vectors of length 5 for all the factors of $w = C(\frac{4}{9})$ of length 9.

While the second order balance matrix of w is given by:

$$U_w = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 & 3 & 2 & 2 & 3 & 2 & 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 & 3 & 2 & 2 & 3 & 2 & 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

with $U_w[5, 5] = 3$.

The value is obtained since $B_w[5] = [0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1]$ and

$$L_5(B_w[5]) = (00110, 01101, 11011, 10110, 01101, 11011, 10110, 01100, 11001, 10010, \\ 00100, 10011)$$

$$|L_5(B_w[5])|_1 = (2, 3, 4, 3, 3, 3, 4, 3, 2, 3, 2, 1, 3)$$

$$U_w[5, 5] = \max(|L_5(B_w[5])|_1) - \min(|L_5(B_w[5])|_1) \\ = 4 - 1 = 3.$$

Therefore $U_w[5, 5] = 3$ due to the factors $|w_1|_1 = |11011|_1 = 4$ and $|w_2|_1 = |00100|_1 = 1$.

Chapter 5

Symmetric standard factorization

At the beginning of the thesis, we gave a geometric definition for the Christoffel words then we introduced an algebraic form. Using the algebraic properties and characteristics, we were able to establish a new order of balancedness for the Christoffel words. But what is the geometric interpretation for this order of balancedness? Where are we able to see it on the Christoffel path? Those questions are answered in this chapter, where a new factorization for the Christoffel words is defined. In Theorem 1.12, we introduced the standard factorization of a Christoffel word w , where for each w , we can find two Christoffel words w_1 and w_2 and write w as $w = w_1w_2$ in a unique way. This factorization is represented by the couple (w_1, w_2) . In order to better understand the geometric interpretation of the second order balance matrix, we define the *Symmetric standard factorization* in section 5.1.

5.1 Symmetric standard factorization

In this section, we define a new tool, named the *symmetric standard factorization*, and used to show a geometric recursive form on the Christoffel words identical to the recursive structure obtained while defining the second order balance matrix in Chapter 3.

Definition 5.1 *Let $w = (w_1, w_2)$ be the standard factorization of the Christoffel word w . In the case where w_1 and w_2 are non-trivial, let w'_1 , w'_2 and w' be the three palindromes such that $w = 0w'1$, $w_1 = 0w'_11$ and $w_2 = 0w'_21$. The symmetric standard factorization of w is the triplet (p, q, p) where p and q are the palindromes such that*

$$p = \begin{cases} w'_1 & \text{if } |w_1| < |w_2|, \\ w'_2 & \text{if } |w_1| > |w_2|. \end{cases} \quad (5.1)$$

and q is the unique word such that: $w = 0pqp1$. This factorization is noted $w = 0(p, q, p)_{ss}1$.

Proposition 5.2 *A Christoffel word $w = (w_1, w_2)$, such that w_1 and w_2 are non-trivial, has a unique symmetric standard factorization $0(p, q, p)_{ss}1$. Moreover p and q are palindromes with $|q| \geq 3$.*

Proof. Let w'_1, w'_2, w' be the three palindromes such that: $w_1 = 0w'_11$; $w_2 = 0w'_21$ and $w = 0w'1$. By the construction of the standard factorization and using the functions ϕ_0 and ϕ_1 , we have $|w_1| = |w_2|$ only if w_1 and w_2 are trivial Christoffel words which is not the case here. Therefore the factor p of the symmetric standard factorization of w is well defined and $2p < |w'|$. Consequently, the factor q exists and both, p and q are palindromes since the central words w'_1 , w'_2 and w' are palindromes. The uniqueness of the standard factorization of w implies the uniqueness of the symmetric standard factorization. Finally:

$$\begin{aligned} |q| &= |w'| - 2|p| \\ &= |w'_1| + |w'_2| + 2 - 2|p| \\ &\geq 3. \end{aligned}$$

□

Example 72 *Let $w = C(5/8) = 0010010100101$ with $(w_1, w_2) = (00100101, 00101)$ its standard factorization. From Theorem 1.12, we know that w_1 and w_2 are Christoffel words and we can write: $w_1 = 0.010010.1 = 0w'_11$ and $w_2 = 0.010.1 = 0w'_21$ where w'_1 and w'_2 are palindromes. We have $|w'_2| < |w'_1|$ then by Definition 5.1, we let $p = w'_2$ and we are able to write:*

$$w = (00100101, 00101) = 0(010, 01010, 010)_{ss}1,$$

where $p = 010$ and $q = 01010$ as we can see in Figure 5.1.

In Chapter 3, we proved that the second order balance is formed of 9 blocks where the blocks α, γ and β are the three main blocks needed for the construction of the matrix. Now in this part, we define a new operator Γ , that is used to construct and identify geometrically those three blocks of $U_{\frac{a}{b}}$. Consider the definition of a Christoffel word given in 2.2 and let (p_0, p_1, \dots, p_n) be the integer points along the path of the Christoffel word w of slope a/b . In the conclusions of Proposition 2.3, we indicated the most important points on the Christoffel path that are beneficial in this thesis. These two points are exactly the closest and the furthest points on

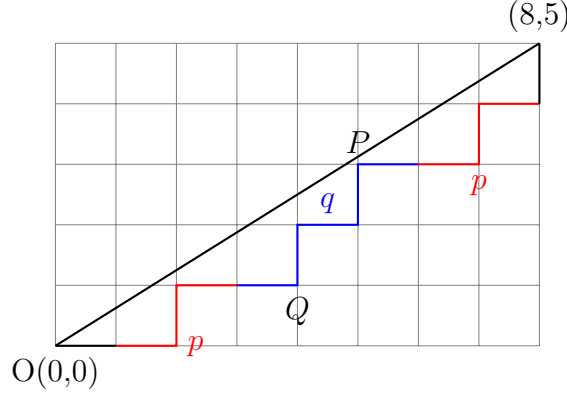


Figure 5.1: The symmetric standard factorization of $w = C(\frac{5}{8})$.

the Christoffel path with respect to the line segment. So we let i be the index of the point that is the closest to the line segment joining $(0, 0)$ to (b, a) , other than $p_0 = (0, 0)$ and $p_n = (b, a)$. In such case, we have $i.a \equiv_n 1$ or equivalently $r[i] = 1$. Similarly, let j be the index of the point that is the furthest from the line segment, then we have $j.a \equiv_n (n - 1)$ or equivalently $r[j] = -1$. The operator Γ extracts the part of a Christoffel word that is between p_i and p_j .

Definition 5.3 Let $w = C(\frac{a}{b})$ and i, j the integers such that $i.a \equiv_n 1$ and $j.a \equiv_n (n - 1)$, where $n = a + b$. The operator Γ is defined as:

$$\Gamma : \mathbb{Q}_+^* \rightarrow A^* \\ \frac{a}{b} \mapsto C\left(\frac{a}{b}\right) [\min(i, j) + 1, \dots, \max(i, j)]$$

The part of w extracted by operator Γ can easily be deduced from its symmetric standard factorization. In order to show this result, we need the following technical lemma.

Lemma 5.4 Let w, p, q be such that $w = 0(p, q, p)_{ss}1$ and let i, j be such that $i.a \equiv_n 1$ and $j.a \equiv_n n - 1$. We have $\{i, j\} = \{|p| + 2, |pq|\}$.

Proof. From Corollary 2.5, we have that $w = (w_1, w_2)$ which implies that $|w_1|.a \equiv_n 1$ and $|w_2|.a \equiv_n (n - 1)$. If $|w_1| < |w_2|$, then by definition $|p| + 2 = |w_1|$ and $|pq| = n - (|p| + 2) = |w_2|$. Conversely, if $|w_1| > |w_2|$ then $|p| + 2 = |w_2|$ and $|pq| = |w_1|$. Finally, the case $|w_1| = |w_2|$ is impossible since it implies that $a/b = 1/1$ and in such case the symmetric standard factorization of w does not exist. \square

Proposition 5.5 *The word $\Gamma(\frac{a}{b})$ is:*

$$\Gamma\left(\frac{a}{b}\right) = \begin{cases} 0^{b-1} & \text{if } a = 1, \\ 1^{a-1} & \text{if } b = 1, \\ q' & \text{where } w = 0(p, q, p)_{ss}1 \text{ and } q = aq'a \text{ for } a \in A, \text{ otherwise.} \end{cases}$$

Proof. Let i and j be two natural numbers such that: $i.a \equiv_n 1$ and $j.a \equiv_n (n-1)$.

- If $a = 1$, then $i = 1$, $j = n - 1$ and $w = C\left(\frac{1}{b}\right) = 0 \cdot 0^{b-1} \cdot 1$.
- If $b = 1$ then $a = n - 1$ so that $i = n - 1$, $j = 1$ and $w = C\left(\frac{a}{1}\right) = 0 \cdot 1^{a-1} \cdot 1$.
- Otherwise, if $a > 1$ and $b > 1$, then the symmetric standard factorization of w exists and, by Proposition 5.2, $|q| \geq 3$. Finally, Lemma 5.4 ensures that $\{i, j\} = \{|p| + 2, |pq|\}$, so that in w , the position $|p| + 2$ is the second letter of q in $0(p, q, p)_{ss}1$ and position $|pq|$ is one letter before the end of q .

□

Note that from Proposition 5.2, $|q| \geq 3$ so that $|p| + 2 < |pq|$ which implies that $\Gamma(\frac{a}{b})$ is not empty. Same thing for the particular cases, while for the fraction $\frac{1}{1}$, we let $\Gamma(\frac{1}{1}) = \epsilon$.

1. If $a = 1$, we have $x = 1$ and $y = n - 1 = b$. Then $\Gamma\left(\frac{1}{b}\right) = C\left(\frac{1}{b}\right) [\min(x, y) + 1 \dots \max(x, y)] = C\left(\frac{1}{b}\right) [x + 1 \dots y] = C\left(\frac{1}{b}\right) [2 \dots b] = 0^{n-2} = 0^{b-1}$.
2. Similarly, if $b = 1$, we have $y = 1$ and $x = n - 1 = a$. Therefore $\Gamma\left(\frac{a}{1}\right) = C\left(\frac{a}{1}\right) [\min(x, y) + 1 \dots \max(x, y)] = C\left(\frac{a}{1}\right) [2 \dots a] = 1^{n-2} = 1^{a-1}$.

Example 73 *Let $w = C(\frac{1}{6})$, by Proposition 5.5, we have that $\Gamma(\frac{1}{6}) = 0^5 1$ referring to the path joining the two critical points of the path, the closest and the furthest point as shown in Figure 5.2.*

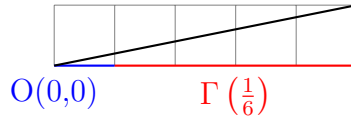


Figure 5.2: The geometrical representation of $\Gamma\left(\frac{1}{6}\right)$ joining the closest and furthest points of the Christoffel path.

As we can see in Figure 5.1, the path joining the two points P and Q , that represent the closest and furthest points of this path with respect to the line segment, is exactly $\Gamma\left(\frac{5}{8}\right)$ and is equal to: 101.

In Chapter 3, we have shown that the second order balance matrix is decomposed into 9 blocks, where mainly 3 of them are sufficient to prove the recursive form obtained. In order to show geometrically this recursive structure, we first need to realize how $C\left(\frac{x}{y}\right)$ is decomposed using the Γ operator. This decomposition is introduced in Lemma 5.6:

Lemma 5.6 *Let $\frac{x}{y} = [a_0, \dots, a_{z-1}]$, $\frac{u}{v} = \text{FEF}\left(\frac{x}{y}\right)$ and $\frac{c}{d} = \text{FDF}\left(\frac{u}{v}\right)$.*

$$C\left(\frac{x}{y}\right) = \begin{cases} \Gamma\left(\frac{u}{v}\right) \Gamma\left(\frac{c}{d}\right) & \text{if } z-1 \text{ is odd,} \\ \Gamma\left(\frac{c}{d}\right) \Gamma\left(\frac{u}{v}\right) & \text{otherwise.} \end{cases}$$

Proof. Recall that we work under the assumption that $x \leq y$ so that the continued fraction of $\frac{x}{y}$ starts with $a_0 = 0$. By definition, we have $\frac{u}{v} = [a_0, \dots, a_{z-1} + 1]$ and $\frac{c}{d} = [a_0, \dots, a_{z-1}, 2]$, There are two cases to consider.

Case 1. $x = 1 \leq y$. In this case, $\frac{x}{y} = [0, y]$ so that $z - 1$ is odd, $\frac{u}{v} = \frac{1}{y+1}$ and $\frac{c}{d} = \frac{2}{2y+1}$. So we can write $C\left(\frac{x}{y}\right) = (0, 0^{y-1}1)_s$ and $C\left(\frac{u}{v}\right) = (0, 0^y1)_s$. By Remark 2.29 (4) we have that $C\left(\frac{c}{d}\right) = (0^{y+1}1, 0^y1)_s$ so that the symmetric standard factorization of $C\left(\frac{c}{d}\right)$ is $0(0^{y-1}, 010, 0^{y-1})_{ss}1$. Finally, Proposition 5.5 states that $\Gamma\left(\frac{u}{v}\right) = 0^y$ and $\Gamma\left(\frac{c}{d}\right) = 1$ which proves the result.

Case 2. $x, y > 1$. in this case, let α, β, p, q be such that $C\left(\frac{x}{y}\right) = (\alpha, \beta)_s = 0(p, q, p)_{ss}1$. In this part, we can get 4 sub-cases depending on the value of $z - 1$ and a_{z-1} :

i If $z - 1$ is odd, from Theorem 2.26 and Lemma 2.38 we have that $\Delta\left(\frac{x}{y}\right)$ ends with letter 0 and $|\alpha| < |\beta|$. The value of a_{z-1} can be equal to 1 or greater or equal to 2, hence we have the following two cases:

1- If $a_{z-1} \geq 2$.

Figure 5.3 illustrates the situation and the proof and we have the following information:

- We have $C\left(\frac{x}{y}\right) = (\alpha, \beta) = 0(p, q, p)_{ss}1$, so that $\alpha = 0p1$ and $\beta = 1^{-1}qp1$.
- Let p', q' be such that $\frac{u}{v} = (\alpha, \alpha\beta)_s = 0(p', q', p')_{ss}1$. We have $p' = p$, $q' = 10pq$ and therefore $\Gamma\left(\frac{u}{v}\right) = 0pq1^{-1}$. Also,

since $p, q, p', q', p'q'p'$ are palindromes, $C(\frac{u}{v}) = 0.p10pqp.1 = 0.pqp01p.1$.

- Let p'', q'' be such that $\frac{c}{d} = (\alpha\alpha\beta, \alpha\beta)_s = 0(p'', q'', p'')_{ss}1$. We have that p'' is the central part of $\alpha\beta$ which is $p'' = pqp$. Using the previous equalities, we have $C(\frac{c}{d}) = \alpha\alpha\beta.\alpha\beta = 0pqp01p1.0pqp1$ so that $q'' = 01p10$. Therefore $\Gamma(\frac{c}{d}) = 1p1$.

Hence, $\Gamma(\frac{u}{v}).\Gamma(\frac{c}{d}) = 0pq1^{-1}.1p1 = 0pqp1 = C(\frac{x}{y})$.

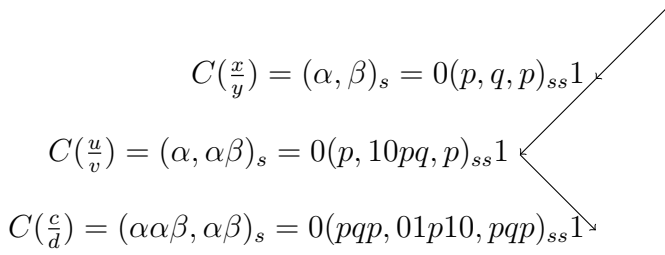


Figure 5.3: The symmetric standard factorization of the Christoffel words of slopes $x/y, u/v$ and c/d with an odd value for a_{z-1} .

2- If $a_{z-1} = 1$.

The relations are presented in Figure 5.4:

- We have $C(\frac{x}{y}) = (\alpha, \beta) = 0(p, q, p)_{ss}1$, so that $\alpha = 0p1$ and $\beta = 1^{-1}qp1$.
- Let p', q' be such that $\frac{u}{v} = (\alpha\beta, \beta) = 0(p', q', p')_{ss}1$. We have $p' = pq, q' = p$. We can write $C(\frac{u}{v}) = 0.pq1^{-1}0^{-1}01p100^{-1}1^{-1}qp1$ since $p, q, p', q', p'q'p'$ are palindromes, and therefore $\Gamma(\frac{u}{v}) = 1p1$.
- Let p'', q'' be such that $\frac{c}{d} = (\alpha\beta, \alpha\beta\beta) = 0(p'', q'', p'')_{ss}1$. We have that p'' is the central part of $\alpha\beta$ which is $p'' = pqp$. Using the previous equalities, we have $C(\frac{c}{d}) = \alpha\beta.\alpha\beta\beta = 0pqp1.0pqpqp1$ so that $q'' = 10pq$. Therefore $\Gamma(\frac{c}{d}) = 0qp1^{-1}$.

Hence, $\Gamma(\frac{c}{d}).\Gamma(\frac{u}{v}) = 0pq1^{-1}.1p1 = 0pqp1 = C(\frac{x}{y})$.

ii If $z - 1$ is even, we have in this case $|\alpha| > |\beta|$ and two sub-cases arise:

3- If $a_{z-1} \geq 2$.

Figure 5.5 illustrates this case:

- We have $C(\frac{x}{y}) = (\alpha, \beta) = 0(p, q, p)_{ss}1$, so that $\alpha = 0pq0^{-1}$ and $\beta = 0p1$.
- Let p', q' be such that $\frac{u}{v} = (\alpha\beta, \beta)_s = 0(p', q', p')_{ss}1$. We have $p' = p, q' = qp10$ and therefore $\Gamma(\frac{u}{v}) = 0^{-1}qpq1$. Also,

$$\begin{array}{c}
C\left(\frac{x}{y}\right) = (\alpha, \beta) = 0(p, q, p)_{ss}1 \\
C\left(\frac{u}{v}\right) = (\alpha\beta, \beta) = 0(0^{-1}1^{-1}qp, 01p10, 0^{-1}1^{-1}qp)_{ss}1 \\
C\left(\frac{c}{d}\right) = (\alpha\beta, \alpha\beta\beta) = 0(pqp, 10pq, pqp)_{ss}1
\end{array}$$

Figure 5.4: The symmetric standard factorization of the Christoffel words of slopes $x/y, u/v$ and c/d with an odd value for a_{z-2} and $a_{z-1} = 2$.

since $p, q, p', q', p'q'p'$ are palindromes, $C\left(\frac{u}{v}\right) = 0.pqp10p.1 = 0.p01pqp.1$.

- Let p'', q'' be such that $\frac{c}{d} = (\alpha\alpha, \alpha\beta\beta)_s = 0(p'', q'', p'')_{ss}1$. We have that p'' is the central part of $\alpha\beta$ which is $p'' = pqp$. Using the previous equalities, we have $C\left(\frac{c}{d}\right) = \alpha\beta.\alpha\beta\beta = 0pqp10p.01pqp1$ so that $q'' = 10p01$. Therefore $\Gamma\left(\frac{c}{d}\right) = 0p0$.

Hence, $\Gamma\left(\frac{c}{d}\right).\Gamma\left(\frac{u}{v}\right) = 0p00^{-1}.qp1 = 0pqp1 = C\left(\frac{x}{y}\right)$.

$$\begin{array}{c}
C\left(\frac{x}{y}\right) = (\alpha, \beta) = 0(p, q, p)_{ss}1 \\
C\left(\frac{u}{v}\right) = (\alpha\beta, \beta) = 0(p, qp10, p)_{ss}1 \\
C\left(\frac{c}{d}\right) = (\alpha\beta, \alpha\beta\beta) = 0(pqp, 10p01, pqp)_{ss}1
\end{array}$$

Figure 5.5: The symmetric standard factorization of the Christoffel words of slopes $x/y, u/v$ and c/d with an even value for a_{z-1} .

3- If $a_{z-1} = 1$.

Figure 5.6 illustrates the situation and we have:

- We have $C\left(\frac{x}{y}\right) = (\alpha, \beta) = 0(p, q, p)_{ss}1$, so that $\alpha = 0pq0^{-1}$ and $\beta = 0p1$.
- Let p', q' be such that $\frac{u}{v} = (\alpha, \alpha\beta) = 0(p', q', p')_{ss}1$. We have $p' = pq0^{-1}1^{-1}$, $q' = 10p01$ and therefore $\Gamma\left(\frac{u}{v}\right) = 0p0$. Also, since $p, q, p', q', p'q'p'$ are palindromes, $C\left(\frac{u}{v}\right) = 0.pq0^{-1}0pqp.1 = 0.pq0^{-1}1^{-1}10p011^{-1}0^{-1}qp.1$.
- Let p'', q'' be such that $\frac{c}{d} = (\alpha\alpha\beta, \alpha\beta) = 0(p'', q'', p'')_{ss}1$. We have that p'' is the central part of $\alpha\alpha\beta$ which is $p'' = pqp$. Using the previous equalities, we have $C\left(\frac{c}{d}\right) = \alpha\alpha\beta.\alpha\beta = 0pqpqp1.0pqp1$ so that $q'' = qp10$. Therefore $\Gamma\left(\frac{c}{d}\right) = 0^{-1}qp1$.

Hence, $\Gamma\left(\frac{u}{v}\right).\Gamma\left(\frac{c}{d}\right) = 0p00^{-1}.qp1 = 0pqp1 = C\left(\frac{x}{y}\right)$.

$$\begin{array}{c}
C\left(\frac{x}{y}\right) = (\alpha, \beta) = 0(p, q, p)_{ss}1 \\
C\left(\frac{u}{v}\right) = (\alpha, \alpha\beta) = 0(pq0^{-1}1^{-1}, 10p01, pq0^{-1}1^{-1})_{ss}1 \\
C\left(\frac{c}{d}\right) = (\alpha\alpha\beta, \alpha\beta) = 0(pqp, qp10, pqp)_{ss}1
\end{array}$$

Figure 5.6: The symmetric standard factorization of the Christoffel words of slopes $x/y, u/v$ and c/d with an even value for a_{z-2} and $a_{z-1} = 2$.

□

After this decomposition, we are able to prove the main result of this section.

Theorem 5.7 *The geometrical form of the γ -block is obtained depending on the position of the fraction in the Stern-Brocot tree and using the Γ operator.*

$$\Gamma\left(\frac{a}{b}\right) = \begin{cases} \Gamma\left(\frac{c}{d}\right) \left[\Gamma\left(\frac{u}{v}\right) \Gamma\left(\frac{c}{d}\right) \right]^{a_z-2} & \text{if } a_z \geq 3 \\ \Gamma\left(\frac{e}{f}\right) & \text{it } a_z = 2 \end{cases} \quad (5.2)$$

where $\frac{e}{f} = \text{FPF}\left(\frac{a}{b}\right)$.

Proof. In this proof, we have 4 distinct cases depending on the position of the fraction in the Stern-Brocot tree and on the morphisms used to reach the fraction a/b .

Let $\frac{a}{b} = [a_0, \dots, a_z]$,

- **Part A: If $a_z \geq 3$:** In this part we can distinguish two cases depending on the morphisms used to reach a/b .

- If we reach a/b with the morphism ϕ_1 , we let p_1, p_2 be two palindromes such that: $C\left(\frac{s}{t}\right) = 0p_11$, $C\left(\frac{g}{h}\right) = 0p_21$ where $C\left(\frac{x}{y}\right) = C\left(\frac{s}{t}\right)C\left(\frac{g}{h}\right)$ and $C\left(\frac{a}{b}\right) = C\left(\frac{s}{t}\right)(C\left(\frac{x}{y}\right))^{a_z}$. Since $p_110p_2 = p_201p_1$, we get:

$$\begin{aligned}
C\left(\frac{a}{b}\right) &= C\left(\frac{s}{t}\right) \cdot \left(C\left(\frac{x}{y}\right) \right)^{a_z} \\
&= 0p_11 \cdot (0p_110p_21)^{a_z-1}0p_110p_21 \\
&= 0p_11 \cdot 0p_201p_11(0p_110p_21)^{a_z-2}0p_110p_21 \\
&= 0(p_110p_2, 01p_11(0p_110p_21)^{a_z-2}0, p_110p_2)_{ss}1.
\end{aligned}$$

Therefore, $\Gamma\left(\frac{a}{b}\right) = 1p_11(0p_110p_21)^{a_z-2}$ and from the first case of Lemma 5.6, we get $\Gamma\left(\frac{a}{b}\right) = \Gamma\left(\frac{c}{d}\right)\left[\Gamma\left(\frac{u}{v}\right)\Gamma\left(\frac{c}{d}\right)\right]^{a_z-2}$.

- If we reach a/b with the morphism ϕ_0 , we let p_1, p_2 be two palindromes such that: $C\left(\frac{s}{t}\right) = 0p_11$, $C\left(\frac{g}{h}\right) = 0p_21$ where $x/y = C\left(\frac{g}{h}\right)C\left(\frac{s}{t}\right)$ and $C\left(\frac{a}{b}\right) = \left(C\left(\frac{x}{y}\right)\right)^{a_z}C\left(\frac{s}{t}\right)$. Since $p_210p_1 = p_101p_2$, we get:

$$\begin{aligned} C\left(\frac{a}{b}\right) &= \left(C\left(\frac{x}{y}\right)\right)^{a_z} \cdot C\left(\frac{s}{t}\right) \\ &= (0p_210p_11)^{a_z-1}0p_210p_11 \cdot 0p_11 \\ &= 0p_210p_11(0p_210p_11)^{a_z-2}0p_101p_210p_11 \\ &= 0(p_210p_1, 1(0p_210p_11)^{a_z-2}0p_101, p_210p_1)_{ss}1. \end{aligned}$$

Therefore, $\Gamma\left(\frac{a}{b}\right) = (0p_210p_11)^{a_z-2}0p_10$ and from the second case of Lemma 5.6, we get:

$$\Gamma\left(\frac{a}{b}\right) = \left[\Gamma\left(\frac{c}{d}\right)\Gamma\left(\frac{u}{v}\right)\right]^{a_z-2}\Gamma\left(\frac{c}{d}\right) = \Gamma\left(\frac{c}{d}\right)\left[\Gamma\left(\frac{u}{v}\right)\Gamma\left(\frac{c}{d}\right)\right]^{a_z-2}.$$

- **Part B: If $a_z = 2$:** Same to the first part, we have two cases depending on the morphisms used to reach a/b :

- With the same givens, if we reach a/b with the morphism ϕ_1 , we have since $p_110p_2 = p_201p_1$:

$$\begin{aligned} C\left(\frac{a}{b}\right) &= C\left(\frac{s}{t}\right) \cdot \left(C\left(\frac{x}{y}\right)\right)^2 \\ &= 0p_11 \cdot 0p_110p_21 \cdot 0p_110p_21 \\ &= 0p_110p_201p_110p_110p_21 \\ &= 0(p_110p_2, 01p_110, p_110p_2)_{ss}1. \end{aligned}$$

Therefore $\Gamma\left(\frac{a}{b}\right) = 1p_11$. We have for the fraction $\frac{e}{f} = \text{FPF}\left(\frac{a}{b}\right)$:

$$\begin{aligned} C\left(\frac{e}{f}\right) &= C\left(\frac{x}{y}\right)C\left(\frac{g}{h}\right) \\ &= 0p_110p_2100p_21 \\ &= 0p_201p_110p_21 \\ &= 0(p_2, 01p_110, p_2)_{ss}1. \end{aligned}$$

Therefore $\Gamma\left(\frac{e}{f}\right) = 1p_11$ and $\Gamma\left(\frac{a}{b}\right) = \Gamma\left(\frac{e}{f}\right)$.

– If we used the morphism ϕ_0 , we have: $p_210p_1 = p_101p_2$ and we get:

$$\begin{aligned} C\left(\frac{a}{b}\right) &= \left(C\left(\frac{x}{y}\right)\right)^2 \cdot C\left(\frac{s}{t}\right) \\ &= 0p_210p_11 \cdot 0p_210p_11 \cdot 0p_11 \\ &= 0p_210p_110p_101p_210p_11 \\ &= 0(p_210p_1, 10p_101, p_210p_1)_{ss}1. \end{aligned}$$

Therefore $\Gamma\left(\frac{a}{b}\right) = 0p_10$. We have that:

$$\begin{aligned} C\left(\frac{e}{f}\right) &= C\left(\frac{g}{h}\right) C\left(\frac{x}{y}\right) \\ &= 0p_210p_210p_11 \\ &= 0p_210p_101p_21 \\ &= 0(p_2, 10p_101, p_2)_{ss}1. \end{aligned}$$

Therefore $\Gamma\left(\frac{e}{f}\right) = 0p_10$ and $\Gamma\left(\frac{a}{b}\right) = \Gamma\left(\frac{e}{f}\right)$.

□

5.1.1 Geometrical examples

At the beginning of this chapter, we mentioned that we will show the geometric link between the blocks of the second order balance matrix $U_{\frac{a}{b}}$ and the Christoffel word. In fact, we present the decomposition of the Christoffel path into several parts where those parts represent exactly the 3 main blocks of the second order balance matrix. We recall that the α -block of $U_{\frac{a}{b}}$ corresponds to the left part of $U_{\text{TBF}(\frac{a}{b})}$, the β -block is exactly the $U_{\text{SUF}(\frac{a}{b})}$ where all the elements of this matrix are increased by 1. Finally the γ -block can be either the γ -block of $\text{FPF}(\frac{a}{b})$ if a/b is at a first deviation position (i.e a/b has the last partial denominator equal to 2), or as presented in equation 3.1, it is the concatenation between the γ -blocks of $\text{TBF}(\frac{a}{b})$ and $\text{FEF}(\text{TBF}(\frac{a}{b}))$.

1. We start with the first case where a/b is at the first deviation position and we pick the fraction $\frac{3}{11}$, where $w = C\left(\frac{3}{11}\right) = 00001000010001$.

From Stern-Brocot tree, we have: $\frac{3}{11} = \frac{1}{4} \oplus \frac{2}{7}$. The second order balance matrix $U_{\frac{3}{11}}$ is constituted of:

- The left part of the second order balance matrix of $\text{TBF}(\frac{3}{11}) = \frac{2}{7}$ that is the α -block of $U_{\frac{3}{11}}$.

$$U_{\frac{3}{11}} = \left(\begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 & 1 & 2 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 & 1 & 2 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \right)$$

- The second order balance matrix of $\text{SUF}(\frac{3}{11}) = \frac{1}{3}$ that is exactly the β -block of $U_{\frac{3}{11}}$ where all the elements of this matrix are increased by 1.

$$U_{\frac{1}{3}} = \left(\begin{array}{c|c|c} 1 & 1 & 1 \\ \hline 1 & 2 & 1 \\ \hline 1 & 1 & 1 \end{array} \right) \xrightarrow{+1} \left(\begin{array}{c|c|c} 2 & 2 & 2 \\ \hline 2 & 3 & 2 \\ \hline 2 & 2 & 2 \end{array} \right)$$

- Since $\frac{3}{11} = [0, 3, 1, 2]$ has a last partial denominator equal to 2 which implies that the rational number is at first deviation position. Hence, the γ -block of $U_{\frac{3}{11}}$ is obtained from the $\text{FPF}(\frac{3}{11}) = \frac{1}{5}$.

$$U_{\frac{1}{5}} = \left(\begin{array}{c|c|c|c|c} 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 2 & 2 & 2 & 1 \\ \hline 1 & 2 & 3 & 2 & 1 \\ \hline 1 & 2 & 2 & 2 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 \end{array} \right)$$

Hence $U_{\frac{3}{11}}$ is given by:

$$U_{\frac{3}{11}} = \left(\begin{array}{ccccc|ccccc|ccccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 2 & 2 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 & 1 & 2 & 3 & 2 & 1 & 1 & 2 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 2 & 2 & 2 & 1 & 2 & 2 & 2 & 1 & 2 & 2 & 2 & 1 & 1 & 1 \\ 1 & 2 & 3 & 2 & 1 & 2 & 3 & 2 & 1 & 2 & 3 & 2 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 1 & 2 & 2 & 2 & 1 & 2 & 2 & 2 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 & 1 & 2 & 3 & 2 & 1 & 1 & 2 & 2 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 2 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \right)$$

In Figure 5.7, we represent the Christoffel word of slope 3/11, we remark the following:

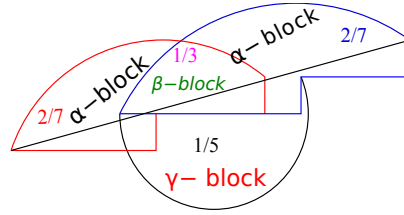


Figure 5.7: The geometric representation of the α , γ and β -blocks of $U_{\frac{3}{11}}$

- 1 Starting from the beginning and the ending of the Christoffel path we are able to find the Christoffel path of slope 2/7 that is needed to obtain the α -block. But we remark that we need to add at the position before the closest point to the line segment a vertical step to get the Christoffel path.
- 2 The intersection of these two paths is exactly the Christoffel path of slope 1/3 referring to the β -block in the second order balance matrix. The extra vertical step in this part is the step responsible of adding +1 to all the entries of the second order matrix of 1/3.
- 3 As shown in this chapter, the Γ operator gives us the path joining the furthest point to the closest point on the Christoffel word with respect to the line segment. Graphically we can see that by adding a horizontal step

at the beginning of this path and a vertical step at the end that refer to 0 and 1, we get the Christoffel path of slope $1/5$ used to obtain the γ -block.

2. The second case concerns the fractions that are with a last partial denominator greater or equal to 3. Which means, the fractions that are in a certain extended branch of the Stern-Brocot tree. We choose the rational number $\frac{6}{11}$ that is equal to: $\frac{6}{11} = \frac{5}{9} \oplus \frac{1}{2} = \frac{1}{1} \oplus \left(\frac{1}{2}\right)^5$ where $w = C\left(\frac{6}{11}\right) = 00100100100100101$. To construct the second order balance matrix of this rational number, exactly like the first case, we need:
 - i For the α -block, we need the left part of $U_{\text{TBF}\left(\frac{6}{11}\right)} = U_{\frac{2}{3}}$;

$$U_{\frac{2}{3}} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$
 - ii For the β -block we need the $U_{\text{SUF}\left(\frac{6}{11}\right)} = U_{\frac{4}{7}}$ where a +1 has to be added to each entry to get the β -block;

$$U_{\frac{4}{7}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 & 2 & 2 & 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 & 2 & 2 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 2 & 2 & 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 & 2 & 2 & 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$
 - iii For the γ -block and as represented in 3.1, we concatenate the gamma part of the fractions $\frac{3}{5}$ and $\frac{2}{3}$.

$$U_{\frac{3}{5}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \quad U_{\frac{2}{3}} = \left(\begin{array}{cc|cc} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{array} \right).$$

Hence the second order matrix has the following form:

$$U_{\frac{6}{11}} = \left(\begin{array}{ccc|cccccccccc|ccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 2 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 & 2 & 2 & 3 & 2 & 2 & 3 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 & 3 & 2 & 3 & 3 & 3 & 3 & 2 & 3 & 2 & 2 & 2 & 2 & 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 & 3 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 & 3 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 & 3 & 2 & 3 & 3 & 3 & 3 & 2 & 3 & 2 & 2 & 2 & 2 & 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 2 & 2 & 3 & 2 & 2 & 3 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 2 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \right).$$

We can see geometrically this decomposition and Figure 5.8 shows the existence of these rational numbers in w where the following remarks can be given:

- (a) In the first case, the intersection between the top branch factor of the Christoffel word from the beginning and at the end is exactly the Christoffel word used for the γ -block. While in this case and since no intersection, we keep on adding the Christoffel word of the top branch fraction and its first extended fraction and we cover the word.
- (b) If we count how many times we added these fractions, we get exactly the same number needed to construct the γ -block of the second order balance matrix.
- (c) Finally, the Christoffel word used for the β -block appears in the middle of the path.

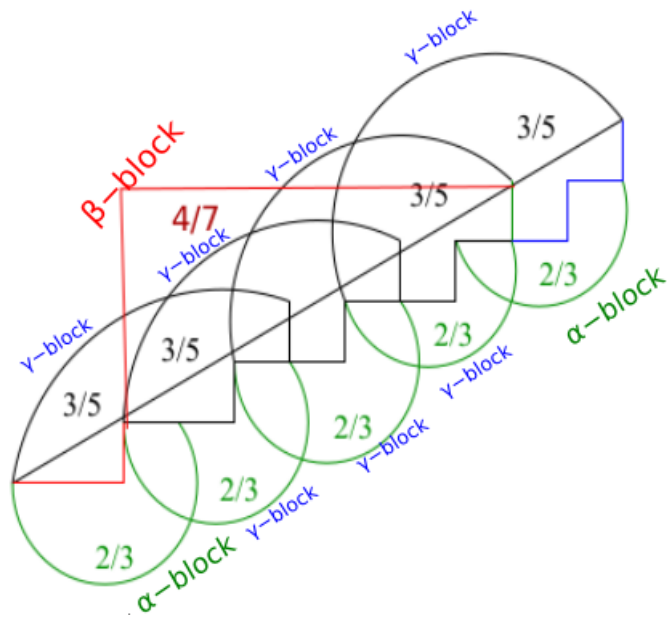


Figure 5.8: The geometric representation of the α , γ and β -blocks of $U_{\frac{6}{11}}$

Chapter 6

Synchronization of Christoffel words

This thesis focuses on Christoffel words and their applications. So far in Chapter 3 and 4, we presented a new order of balancedness and we did the link with the Abelian vectors. Christoffel words are one of the special types of words in combinatorics that have a geometrical interpretation as we have already seen. Many studies were done by Elwin Christoffel [25], Borel [7], Berstel [4] and Reutenauer [5] to find all the characterization of these words. Finding balanced words was always the purpose of lot of mathematicians like Etian Altman [2], Pascal Hubert [32] and Robert Tijdeman [51]. The usual Christoffel word can be seen as a synchronization using two generators like the work done by Reutenauer [38]. Aviezri Fraenkel, in his papers [24],[25], [50, 52] has mentioned many ways to prove that the unique solution for balanced words on k -letters alphabet with pairwise distinct frequencies of letter, where $3 \leq k \leq 7$, is the sequence of generators of the form: 2^{k-1} called **Fraenkel's conjecture** for all $k \geq 3$. Many studies use either balanced words or Beatty sequences like Ronald Graham [30], nevertheless Christoffel words can give a better insight on Fraenkel's conjecture.

In this chapter, we present how two Christoffel words are synchronized and how three Christoffel words can be synchronized. In the first section we define some invariants that help us to find a simple way to write the synchronization of these Christoffel words using a specific seed. In the second section we find a relation between the synchronization of two and three Christoffel words, hence we are able to pass simply from one case to another.

6.1 Synchronization of Christoffel words

In the thesis of G.Paquin, we can find an interpretation for the synchronization of k Christoffel words on k letters and an extra symbol $*$ and many properties of partial synchronization of Christoffel words (the $*$ materialize an extra gap letter in the synchronization). Many properties from the Beatty sequences point of view to the combinatorics on words point of view [38] were used. Now we use explicitly the properties of Christoffel words in order to make the synchronization of 3 Christoffel words. The concepts developed will certainly help to attack the Fraenkel's conjecture as we will see at the end of this chapter.

Given d well chosen Christoffel words, written respectively on alphabets $\{1, *\}$, $\{2, *\}$, \dots , $\{d, *\}$, a **synchronization** provides a word on alphabet $\{1, 2, \dots, d\}$. This notion is closely related to the **superimposition** defined in [38] with the difference that we do not only require the words to be *collision free* on the letter $\{1, 2, \dots, d\}$, we also require that there is a letter among $\{1, 2, \dots, d\}$ at each position. That is to say that the final word does not contain any occurrences of the letter $*$. More formally, let g_1, g_2, \dots, g_d be strictly positive integers called **generators** and s_1, s_2, \dots, s_d be integers called **seeds**. In order to lighten the presentation, the vector $G = (g_1, \dots, g_d)$ is called a **generator** and $S = (s_1, s_2, \dots, s_d)$ is called a **seed**. We denote $P_a(g, n, s)$ the word obtained from the Graph of Cayley of the generator g modulo n starting with the seed s over the alphabet $A = \{a, *\}$.

Definition 6.1 *Given generators g_1, g_2, \dots, g_d , and seeds s_1, s_2, \dots, s_d , let $n = g_1 + \dots + g_d$. A superposition of Christoffel words is given by the two matrices $V = (v_{i,j})_{\substack{i=1..d \\ j=0..n}}$ and $E = (e_{i,j})_{\substack{i=1..d \\ j=0..n-1}}$, defined as:*

$$v_{i,j} = (s_i + jg_i) \pmod n \quad \text{and} \quad e_{i,j} = \begin{cases} i & \text{if } v_{i,j} > v_{i,j+1} \\ * & \text{otherwise} \end{cases}$$

We remark that: $e_i = P_i(g_i, n, s_i)$.

Informally, a superposition of Christoffel words, is equivalent to the draw of Cayley graphs of the Christoffel words of slope $g_i/(n - g_i)$ linearly, one below the other.

Example 74 *For the generators $g_1 = 5$, $g_2 = 3$ and the seeds $s_1 = 0$, $s_2 = 0$, we can write:*

$$\begin{aligned} g_1 = 5 : & 0 \xrightarrow{*} 5 \xrightarrow{1} 2 \xrightarrow{*} 7 \xrightarrow{1} 4 \xrightarrow{1} 1 \xrightarrow{*} 6 \xrightarrow{1} 3 \xrightarrow{1} 0 \\ g_2 = 3 : & 0 \xrightarrow{*} 3 \xrightarrow{*} 6 \xrightarrow{2} 1 \xrightarrow{*} 4 \xrightarrow{*} 7 \xrightarrow{2} 2 \xrightarrow{*} 5 \xrightarrow{2} 0 \end{aligned}$$

In this example, we got a matrix that contains the values of the vertices of the Cayley graphs of each generator and a second matrix that represents the edges. In order to get the matrix of synchronization V and the matrix of the edges E , we must introduce the following definition:

Definition 6.2 *The seed S **synchronizes** the generator G if each column of the matrix E contains exactly $d - 1$ occurrences of the letter $*$ and one occurrence of a letter in $\{1, 2, \dots, d\}$. We call the synchronization word the word obtain by the concatenation of the letters in $\{1, 2, \dots, d\}$.*

Example 75 *The superposition for generators $g_1 = 5$, $g_2 = 3$ and seeds $s_1 = 0$, $s_2 = 7$ form a synchronization :*

$$\begin{array}{cccccccc} 0 & \xrightarrow{*} & 5 & \xrightarrow{1} & 2 & \xrightarrow{*} & 7 & \xrightarrow{1} & 4 & \xrightarrow{1} & 1 & \xrightarrow{*} & 6 & \xrightarrow{1} & 3 & \xrightarrow{1} & 0 \\ 7 & \xrightarrow{2} & 2 & \xrightarrow{*} & 5 & \xrightarrow{2} & 0 & \xrightarrow{*} & 3 & \xrightarrow{*} & 6 & \xrightarrow{2} & 1 & \xrightarrow{*} & 4 & \xrightarrow{*} & 7 \end{array}$$

The synchronization word is 21211211, and the seed is $\begin{pmatrix} 0 \\ 7 \end{pmatrix}$. The two matrices V and E are given as the following:

$$V = \begin{pmatrix} 0 & 5 & 2 & 7 & 4 & 1 & 6 & 3 & 0 \\ 7 & 2 & 5 & 0 & 3 & 6 & 1 & 4 & 7 \end{pmatrix} \text{ and } E = \begin{pmatrix} * & 1 & * & 1 & 1 & * & 1 & 1 \\ 2 & * & 2 & * & * & 2 & * & * \end{pmatrix}.$$

The geometrical representation can be given by:

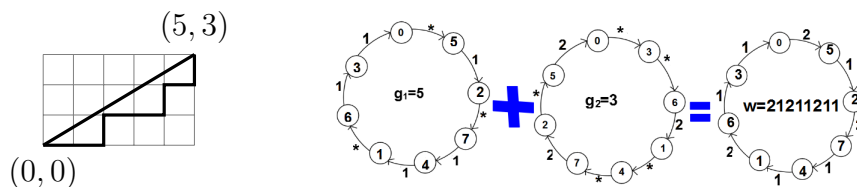


Figure 6.1: Two representation of the Christoffel word 11211212 of slope 3/5. On the left as a lattice path and on the right as a Cayley graphs. The representation as Cayley graphs highlights the fact that the Christoffel word of slope p/q has the same structure then the one of slope q/p and that both synchronizes into a circular conjugate.

The aim of this chapter is to provide a complete characterization of the seeds that provides a synchronization for any triplet of generators g_1, g_2, g_3 .

6.2 Vertical invariant

For instance, we are going to define the vertical and horizontal invariants in order to study and prove that any three generators that are not pairwise distinct can be synchronized and to find their seeds.

Definition 6.3 *Let $w \in A^*$ be a balanced word over a k -letter alphabet, with $k \geq 3$ such that w is the synchronization word obtained from the superposition of k generators. We let V the proper matrix of synchronization of these generators and we define $c_j = \sum_{i=1}^d v_{i,j} \forall j$.*

Theorem 6.4 *The c_j 's are constant if and only if there exists a unique decreasing position in each column of the synchronization matrix starting from a specific seed.*

Proof. \Rightarrow By contradiction, we let V be the superposition matrix of Christoffel words defined like above. We choose a column j and we let c_j be the constant value obtained on each column. We have $c_j = \sum_{i=1}^d v_{i,j} = u$ and we let all the elements of the next column be greater than the $v_{i,j}$'s but smaller than n , in this case all the elements of $e_{i,j}$ are $*$, which means there are any decreasing in this column.

$$\begin{aligned} v_{1,j+1} &= v_{1,j} + g_1 < n \\ v_{2,j+1} &= v_{2,j} + g_2 < n \\ &\vdots \\ v_{d,j+1} &= v_{d,j} + g_d < n \end{aligned}$$

The sum of the elements of the column $j + 1$ is equal to: $c_{j+1} = \sum_{i=1}^d v_{i,j+1} =$

$$\sum_{i=1}^d v_{i,j} + g_1 + g_2 + \dots + g_d = c_j + n = u + n \neq u.$$

Contradiction since the constant is not respected on each column.

Suppose now that we have two decreasing positions at row k and l such that $k < l$, respecting all the previous conditions. We get the following:

$$\begin{aligned}
v_{1,j+1} &= v_{1,j} + g_1 < n \\
v_{2,j+1} &= v_{2,j} + g_2 < n \\
&\vdots \\
v_{k,j+1} &= v_{k,j} + g_k > n = v_{k,j} + g_k - n \\
&\vdots \\
v_{l,j+1} &= v_{l,j} + g_l > n = v_{l,j} + g_l - n \\
&\vdots \\
v_{d,j+1} &= v_{d,j} + g_d < n
\end{aligned}$$

Then the elements of $e_{i,j}$ are all $*$ unless $e_{k,j} = k$ and $e_{l,j} = l$. By calculating

$$c_{j+1} = \sum_{i=1}^d v_{i,j+1} = \sum_{i=1}^d v_{i,j} + g_1 + g_2 + \dots + g_d - n - n = c_j + n - 2n = u - n \neq u.$$

Hence the vertical invariant is not respected too when we have more than one decreasing. Same result is obtained if we have s decreasing instead of two. The $c_{i,j+1} = u - (s - 1)n$ which is a contradiction.

\Leftarrow we fix a column j and we say:

$$\begin{aligned}
s_{1,j} &= s_{1,0} + jg_1 \\
&\vdots \\
s_{d,j} &= s_{d,0} + jg_d
\end{aligned}$$

hence $c_j = c_0 + jn$ and since by hypothesis we know that we have exactly one decreasing from a column to another then to reach the j^{th} column, we have to remove n j times and then we have to subtract jn from c_j and hence we get: $c_j = c_0 = u \forall j \in \{0, \dots, n\}$. \square

NOTATION: We call this constant the Vertical invariant and we denote it by $I_{n,k}$.

Lemma 6.5 Let $d_i = \gcd(g_i, n)$. Then $k \leq \sum_{i=1}^k d_i \leq n$.

Proof. By its definition, $1 \leq d_i \leq g_i$. Thus, $k \leq \sum_{i=1}^k d_i \leq n$. \square

Proposition 6.6 $I_{n,k} = \frac{kn}{2} - \frac{1}{2} \sum_{i=1}^k d_i$, where $d_i = \gcd(g_i, n)$.

Proof. Let g_i be the generator of the i -th row and let $d_i = \gcd(g_i, n)$. Then, the $\pi_i(w)$ is a conjugate of a d_i -th power of the Christoffel word of length n/d_i having g_i/d_i occurrences of the smallest letter. Thus, the values on the i -th row are the ones of d_i times the Cayley graph of the word $C(n/d_i, g_i/d_i)$ generated by $d_i \cdot g_i/d_i$. The sum of the values of the i -th rows is $d_i \sum_{j=1}^{n/d_i-1} j d_i$ and so, the sum of the k -th rows is $\sum_{i=1}^k \left(d_i \sum_{j=1}^{n/d_i-1} j d_i \right)$. Since the sum of each of the n columns are equal, we obtain

$$\begin{aligned} I_{n,k} &= \frac{\sum_{i=1}^k \left(d_i \sum_{j=1}^{n/d_i-1} j d_i \right)}{n} = \frac{\sum_{i=1}^k \left(d_i^2 \sum_{j=1}^{n/d_i-1} j \right)}{n} = \frac{1}{n} \sum_{i=1}^k d_i^2 \frac{\left(\frac{n}{d_i} - 1 \right) \frac{n}{d_i}}{2} \\ &= \frac{1}{n} \sum_{i=1}^k \frac{(n^2 - n d_i)}{2} = \frac{k n}{2} - \frac{1}{2} \sum_{i=1}^k d_i. \end{aligned}$$

□

Lemma 6.7 For $(g_1, g_2) \bmod n$ where $n = g_1 + g_2$ with $d = \gcd(g_i, n) \forall i \in \{1, 2\}$, we get $I_{n,k} = n - d$

Proof. Let $(g_1, g_2) \bmod n$ where $n = g_1 + g_2$ with $d = \gcd(g_i, n) \forall i \in \{1, 2\}$, we get:

$$I_{n,k} = \frac{k \cdot n}{2} - \frac{1}{2} \left(\sum_{i=1}^2 \gcd(n, g_i) \right) = \frac{2n}{2} - \frac{1}{2}(d + d) = n - d,$$

□

Example 76 From Example 75, we presented the synchronization of the generators $g_1 = 3$ and $g_2 = 5$. If we compare the sum of each column obtained after synchronization, we can notice that it is a constant value equal to: $n - 1 = 7$.

6.3 Seeds for two generators

Geneviève Paquin and Reutenauer in their paper [38] have proved that two generators are always synchronized for all n in particular when n is equal to their sum. In addition to that, since we know the value of $I_{n,k} = n - d$, then we can deduce that the seed for two generators is always of the form:

$$S = \begin{pmatrix} 0 \\ n - d \end{pmatrix} = \begin{pmatrix} 0 \\ I_{n,k} \end{pmatrix}$$

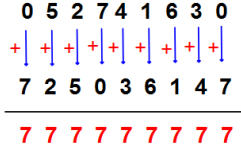


Figure 6.2: The vertical invariant of the synchronized form of the generators $g_1 = 3$ and $g_2 = 5$.

Example 77 Let $(g_1, g_2) = (2, 4) \pmod 6$, where $I_{n,k} = I_{6,k} = n - d = 6 - 2 = 4$, hence we get the following synchronization:

$$\begin{aligned}
 g_1 &= 2 : 0 \xrightarrow{*} 2 \xrightarrow{*} 4 \xrightarrow{1} 0 \xrightarrow{*} 2 \xrightarrow{*} 4 \xrightarrow{1} 0 \\
 g_2 &= 4 : 4 \xrightarrow{2} 2 \xrightarrow{2} 0 \xrightarrow{*} 4 \xrightarrow{2} 2 \xrightarrow{2} 0 \xrightarrow{*} 4
 \end{aligned}$$

where the word formed is: $w = 221221$ and the seed is: $S = \begin{pmatrix} 0 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ n - d \end{pmatrix}$.

In Chapter 2, we defined the set of decreasing positions of a Christoffel word of slope a/b in Theorem 2.8 by $D(a, b)$. This set contains all the positions of the letter 1 of the word, where the positions are counted from 1 to $n = a + b$. In fact, as we mentioned before, Paquin and Reutenauer in their paper [38], corollary 3.2 have defined $C'(n, g, 0) \forall n$ where $\gcd(g, n) = 1$ as being the set of all increasing positions of the generator $g \pmod n$ starting by 0. It is obtained as follows: $C'(n, g, 0) = \{i\alpha'; i = 0 \dots n - g - 1\}$ where $\alpha' \cdot (n - g) \equiv -1 \pmod n$. This result remains correct in particular where $n = \sum g_i, \forall i$ like in the case of Christoffel words. Therefore, we denote the set of decreasing positions by: $D(n, g, 0)$, where in particular $D(a, b) = D(a + b, a, 1)$. This definition is correct when g and n are coprime and the position of the word starts at 0. The following lemma is deduced from this theorem and determines the form of the set of decreasing positions of a word w of slope a/b with a and b not coprime, and where the word starts at position 0.

Lemma 6.8 Let $a = a'd, b = b'd$ and $n = n'd$ where $d = \gcd(a, b)$, the set of decreasing positions for w is the set:

$$D(n'd, a'd, 0) = \cup_{j=0}^{d-1} \{i\alpha; i = 1, \dots, a' + jn'\} \text{ where } \alpha \cdot a \equiv -d \pmod n.$$

Proof. We let $g_1 = n - g$ then $d = \gcd(g, n) = \gcd(g_1, n)$, $\frac{g_1}{d} \perp \frac{n}{d}$ and $\gcd(\frac{g_1}{d}, \frac{n}{d}) = 1$. We also know that any two generators are synchronized with seed $\begin{pmatrix} 0 \\ n - d \end{pmatrix}$ having the increasing positions of g correspond exactly to the decreasing positions of g_1 by Remark 1.8. We let $D(n, g, 0)$ be the set of decreasing positions for g and $C(n, g_1, n - d)$ be the set of increasing of g_1 starting by $n - d$ using [38]. $C(n, g_1, n - d) = \{i\alpha' + x; i = 0, \dots, (\frac{n-g_1}{d}) - 1\}$; $\frac{n-g_1}{d}.\alpha' \equiv -1 \pmod{n'}$ and x being the number of shifts needed to start from $n - d$. Hence $C(n, g_1, n - d) = \{i\alpha' + x; i = 0, \dots, g' - 1\}$; $g'.\alpha' \equiv -1 \pmod{n'}$ which is equivalent to say that: $g.\alpha' \equiv -d \pmod{n}$ since:

$$\begin{aligned} g'.\alpha' &\equiv -1 \pmod{n'} \\ g'.\alpha' &= n'k - 1; k \in \mathbb{Z} \\ dg'.\alpha' &= n'kd - d \\ g.\alpha' &= nk - d \\ g.\alpha' &\equiv -d \pmod{n}. \end{aligned}$$

The step x that must be used to start from $n - d$ instead of 0 is calculated as follows:

$$\begin{aligned} n - d + xg_1 &\equiv 0 \pmod{n} \\ xg_1 &\equiv d \pmod{n} \\ nx - xg &\equiv d \pmod{n} \\ xg &\equiv -d \pmod{n} \end{aligned}$$

hence $x = \alpha'$ and the set $C(n, g_1, n - d) = \{(i + 1)\alpha'; i = 0 \dots g' - 1\} = \{i\alpha'; i = 1, \dots, g'\}$, where $\alpha'.g \equiv -d$. □

In Chapter 1, we introduced the circular permutation of a word at a position i and we said that $\sigma_i(w) = \sigma_i(a_1 \dots a_n) = a_{i+1} \dots a_n a_1 \dots a_i$. In particular and for a letter a in the alphabet A and a finite word $w \in A^*$, the conjugacy operator μ is defined in [38] by $\mu(aw) = wa$.

Proposition 6.9 $P_a(g, n, s) = \mu^x P_a(g, n, 0)$ where $xg \equiv -s \pmod{n}$ over the alphabet $A = \{a, *\}$.

Proof. x represents the number of circular permutations we need to pass from s to 0. Since the operator μ represents a circular permutation then we need to compose μ , x times, hence we have: $P(g, n, s) = \mu^x P(g, n, 0)$ with:

$$\begin{aligned} s_i + x.g &\equiv 0 \pmod{n} \\ x.g &\equiv -s_i \pmod{n} \end{aligned}$$

□

Example 78 *Synchronization form with three generators*

Let the triplet $G = (1, 2, 4) \pmod{7}$.

$$\begin{aligned} g_1 &= 1 : 4 \xrightarrow{*} 5 \xrightarrow{*} 6 \xrightarrow{1} 0 \xrightarrow{*} 1 \xrightarrow{*} 2 \xrightarrow{*} 3 \xrightarrow{*} 4 \\ g_2 &= 2 : 5 \xrightarrow{2} 0 \xrightarrow{*} 2 \xrightarrow{*} 4 \xrightarrow{*} 6 \xrightarrow{2} 1 \xrightarrow{*} 3 \xrightarrow{*} 5 \\ g_3 &= 4 : 0 \xrightarrow{*} 4 \xrightarrow{3} 1 \xrightarrow{*} 5 \xrightarrow{3} 2 \xrightarrow{*} 6 \xrightarrow{3} 3 \xrightarrow{3} 0 \end{aligned}$$

We can note that: $I_{n,k} = 9$.

6.4 Horizontal invariant

$P(g, n, 0)$ is a Christoffel word made of the product of two palindromes. As we have defined before, $P(g, n, s) = \mu^x P(g, n, 0)$, then it remains the product of two palindromes. In this case we have one center of symmetry for each palindrome, hence two centers of symmetries for each word that we can calculate its positions using the following relations:

$$x_{c_1} = \frac{l-1}{2}; \quad x_{c_2} = \frac{l+n-1}{2} \quad \text{where } g.l \equiv n-d \pmod{n},$$

with $\mu^x(x_{c_i}) = x_{c_i} + x$; where $x.g \equiv -s \pmod{n} \forall i \in \{1, 2\}$.

Definition 6.10 We define the **Horizontal invariant** to be the constant value $I_{h_{g_i}}$ obtained when we add any two equidistant values from the center of symmetry.

Lemma 6.11 Let k be the number of generators to study and $n = g_1 + g_2 + \dots + g_k$. The horizontal invariant $I_{h_{g_i}} = n - d_i \forall i \in \{1, \dots, k\}$, where $d_i = \gcd(n, g_i)$.

Proof. From the definition,

$$\begin{aligned} I_{h_{g_i}} &= x_{c_1}g + (x_{c_1} + 1)g \\ I_{h_{g_i}} &= \frac{l-1}{2}g + \left(\frac{l-1}{2} + 1\right)g \\ I_{h_{g_i}} &= lg \equiv n - d_i \pmod{n}. \end{aligned}$$

Or:

$$\begin{aligned} I_{h_{g_i}} &= x_{c_2}g + (x_{c_2} + 1)g \\ I_{h_{g_i}} &= \frac{l+n-1}{2}g + \left(\frac{l+n-1}{2} + 1\right)g \\ I_{h_{g_i}} &= (l+n)g \\ I_{h_{g_i}} &= lg + ng \equiv n - d_i \pmod{n}. \end{aligned}$$

□

Example 79 Let $g = 3$ and $n = 7$ then the orbits of g in $\mathbb{Z}/7\mathbb{Z}$ are:

$$0 \xrightarrow{*} 3 \xrightarrow{*} 6 \xrightarrow{1} 2 \xrightarrow{*} 5 \xrightarrow{1} 1 \xrightarrow{*} 4 \xrightarrow{1} 0, \text{ where } I_{h_3} = n - 1 = 6.$$

For $g = 2$ and $n = 6$, the orbits of g in $\mathbb{Z}/6\mathbb{Z}$ are:

$$0 \xrightarrow{*} 2 \xrightarrow{*} 4 \xrightarrow{1} 0 \xrightarrow{*} 2 \xrightarrow{*} 4 \xrightarrow{1} 0, \text{ where } I_{h_2} = n - d = n - 2 = 4.$$

We can note that in this case $I_{h_{g_i}} = d \cdot I_{h_{\frac{g_i}{d}}}$.

6.5 Case of three generators

After defining the sets $D(g, n, s)$ and $C(g, n, s)$, we introduce now the seeds for the case of $k = 3$ and study the synchronization of three generators. Hence, we let the triplet $G = (g_1, g_2, g_3)$ with $n = g_1 + g_2 + g_3$ and we start by giving some relations and algebraic properties for some particular cases. We let two of the generators to have the same value like for example: $g_1 = g_2$; $g_1 = g_3$ or $g_2 = g_3$ with $\gcd(g_i, g_j) = 1, \forall i, j \in \{1, 2, 3\}; i \neq j$. For simplification, we assume that $g_1 = g_2 = g$, and g_3 is distinct from g , since the other cases give us exactly the same result.

Proposition 6.12 For the triplet $G = (g, g, g_3)$ with $n = 2g + g_3$, n is an odd number if and only if $\gcd(n, g) = \gcd(n, g_3) = 1$.

Proof. “ \Leftarrow ” For $G = (g, g, g_3)$ and $n = 2g + g_3$, we can write $n = 2k$ if n is an even number, and in this case g_3 has to be an even number too since $2g$ is even. Therefore $\gcd(n, g_3) = \gcd(n, 2g) = 2$, contradiction.

“ \Rightarrow ” By the **Euclidean Algorithm**, $\gcd(g, g_3) = 1$ implies $\gcd(n, g) = \gcd(g, g_3) = 1$. Let $d = \gcd(n, g_3)$, then $d|n$ and $d|g_3$, hence $d|(2g + g_3)$. Therefore, there exists $\alpha \in \mathbb{N}$; $2g + g_3 = \alpha d$ and $\exists \beta \in \mathbb{N}$; $g_3 = \beta d$. Hence $2g = (\alpha - \beta)d$ and $d|2g$.

Since n is odd and $n = 2g + g_3$ therefore g_3 is odd and $d|g_3$, so we get d odd. $d|2g$ then $\exists k \in \mathbb{N}$; $2g = kd$. Since $2g$ is even and d is odd then k is even and we can write $g = \frac{k}{2}d$, hence $d|g$. Now $d|g$; $d|n$ and $\gcd(n, g) = 1$ therefore $d = 1$. \square

Lemma 6.13 For the triplet $G = (g, g, g_3)$ and $n = 2g + g_3$, if n is an odd number then:

1. $I_{n,k} = \frac{3}{2}(n - 1)$.

2. $I_{h_g} = I_{h_{g_3}} = n - 1$.

Proof. Since n is an odd number then $\gcd(n, g) = \gcd(n, g_3) = 1$

$$I_{n,k} = \frac{k \cdot n}{2} - \frac{1}{2} \left(\sum_{i=1}^3 \gcd(n, g_i) \right) = \frac{3n}{2} - \frac{1}{2} \left(\sum_{i=1}^3 1 \right) = \frac{3n}{2} - \frac{3}{2} = \frac{3}{2}(n - 1).$$

With this, we can conclude that the horizontal invariant is: $I_{h_g} = I_{h_{g_3}} = n - 1$ \square

Proposition 6.14 For the triplet $G(g, g, g_3)$ with $n = 2g + g_3$, n is an even value if and only if $\gcd(n, g) = 1$ and $\gcd(n, g_3) = 2$.

Proof. “ \Rightarrow ” Since n is even and $n = 2g + g_3$ then g_3 is even.

$\gcd(g, g_3) = 1$ implies g is odd and $\gcd(n, g) = 1$ by the Euclidean Algorithm.

g_3 is even and n is even, hence $2|g_3$ and $2|n$.

Let $d \neq 1$ be another divisor of g_3 and n ; $d|g_3$ and $d|n$, then $\exists \alpha \in \mathbb{N}$; $2g + g_3 = \alpha d$ and $\exists \beta \in \mathbb{N}$; $g_3 = \beta d$. Therefore, $2g = (\alpha - \beta)d$ and $d|2g$.

If $d \nmid 2$ then $d|g$, contradiction, since $d|g$ and $d|n$ with $\gcd(n, g) = 1$ hence $d = 1$ contradiction.

Therefore, $d|2 \Rightarrow \exists \alpha' \in \mathbb{N}$; $2 = \alpha'd$. and hence $d = 2$ or $d = 1$ and since by hypothesis $d \neq 1$ then certainly $d = 2$.

“ \Leftarrow ” If $\gcd(n, g) = 1$ and $\gcd(n, g_3) = 2$ then $2|n; 2|g_3$ hence g_3 is even. $n = 2g + g_3$; $2g$ is even, g_3 is even, then n is even. \square

Lemma 6.15 For the triplet $G = (g, g, g_3)$ and $n = 2g + g_3$, if n is an even value then:

1. $I_{n,k} = \frac{3}{2}n - 2 = \frac{3n-4}{2}$.
2. $I_{h_g} = n - 1$ and $I_{h_{g_3}} = n - 2$.

Proof. Since n is even then $\gcd(n, g) = 1$ and $\gcd(n, g_3) = 2$.

$$I_{n,k} = \frac{k.n}{2} - \frac{1}{2} \left(\sum_{i=1}^3 \gcd(n, g_i) \right) = \frac{3n}{2} - \frac{1}{2}(1 + 1 + 2) = \frac{3n - 4}{2} = \frac{3}{2}n - 2.$$

In addition, with n as an even value and $I_{h_{g_i}} = n - d$ then $\gcd(n, g) = 1$, $\gcd(n, g_3) = 2$ and $I_{h_g} = n - 1$, $I_{h_{g_3}} = n - 2$. \square

Example $G = (3, 3, 4)$

Let $(g_1, g_2, g_3) = (3, 3, 4)$. We have $n = g_1 + g_2 + g_3 = 10$. Orbits in $\mathbb{Z}/10\mathbb{Z}$ of generators g_1, g_2 and g_3 are represented as rows below:

$$\begin{aligned} g_1 &= 3 : 4 \xrightarrow{*} 7 \xrightarrow{1} 0 \xrightarrow{*} 3 \xrightarrow{*} 6 \xrightarrow{*} 9 \xrightarrow{1} 2 \xrightarrow{*} 5 \xrightarrow{*} 8 \xrightarrow{1} 1 \xrightarrow{*} 4 \\ g_2 &= 3 : 9 \xrightarrow{2} 2 \xrightarrow{*} 5 \xrightarrow{*} 8 \xrightarrow{2} 1 \xrightarrow{*} 4 \xrightarrow{*} 7 \xrightarrow{2} 0 \xrightarrow{*} 3 \xrightarrow{*} 6 \xrightarrow{*} 9 \\ g_3 &= 4 : 0 \xrightarrow{*} 4 \xrightarrow{*} 8 \xrightarrow{3} 2 \xrightarrow{*} 6 \xrightarrow{3} 0 \xrightarrow{*} 4 \xrightarrow{*} 8 \xrightarrow{3} 2 \xrightarrow{*} 6 \xrightarrow{3} 0 \end{aligned}$$

In this example we can see that the $I_{n,k} = \frac{3.n-4}{2} = \frac{3.10-4}{2} = 13$ since n is an even number, $I_{h_4} = 8 = n - d = n - 2$ and $I_{h_3} = 9 = n - 1$. The word formed is:

$w = 2132312313$ with $S = \begin{pmatrix} 4 \\ 9 \\ 0 \end{pmatrix}$ its seed.

Example $G = (4, 4, 3)$

Let $(g_1, g_2, g_3) = (4, 4, 3)$. We have $n = g_1 + g_2 + g_3 = 11$. Orbits in $\mathbb{Z}/11\mathbb{Z}$ of generators g_1, g_2 and g_3 are represented as rows below:

$$\begin{aligned}
g_1 &= 4 : 5 \xrightarrow{*} 9 \xrightarrow{1} 2 \xrightarrow{*} 6 \xrightarrow{*} 10 \xrightarrow{1} 3 \xrightarrow{*} 7 \xrightarrow{1} 0 \xrightarrow{*} 4 \xrightarrow{*} 8 \xrightarrow{1} 1 \xrightarrow{*} 5 \\
g_2 &= 4 : 10 \xrightarrow{2} 3 \xrightarrow{*} 7 \xrightarrow{2} 0 \xrightarrow{*} 4 \xrightarrow{*} 8 \xrightarrow{2} 1 \xrightarrow{*} 5 \xrightarrow{*} 9 \xrightarrow{2} 2 \xrightarrow{*} 6 \xrightarrow{*} 10 \\
g_3 &= 3 : 0 \xrightarrow{*} 3 \xrightarrow{*} 6 \xrightarrow{*} 9 \xrightarrow{3} 1 \xrightarrow{*} 4 \xrightarrow{*} 7 \xrightarrow{*} 10 \xrightarrow{3} 2 \xrightarrow{*} 5 \xrightarrow{*} 8 \xrightarrow{3} 0
\end{aligned}$$

In this example we can see that the $I_{n,k} = \frac{3.n-3}{2} = \frac{3.11-3}{2} = 15$ since n is an odd number, $I_{h_4} = I_{h_3} = 10 = n - 1$ and $w = 21231213213$ is the word formed with $S = \begin{pmatrix} 5 \\ 10 \\ 0 \end{pmatrix}$ its seed.

6.5.1 Generators not relatively co prime

In this part we consider that the generators have a common divisor d and we define the invariants and the seeds in this case.

Lemma 6.16 *Let $\gcd(g_i, g_j) = d$ then $I_{n,k} = d.I_{\frac{n}{d},k}$ and $I_{h_{g_i}} = n - d$ or $n - 2d$.*

Proof. Since $\gcd(g, g_3) = d$ then $d|g; d|g_3$ hence d divides any combination between g and g_3 , in particular $2g + g_3 = n$, hence $d|n$. By dividing by d we can get $(\frac{g}{d}, \frac{g}{d}, \frac{g_3}{d}) \pmod{\frac{n}{d}}$ with $\gcd(\frac{g}{d}, \frac{g_3}{d}) = 1$.

Using Propositions 6.12 and Proposition 6.14 we get:

- If $\frac{n}{d}$ is odd then $I_{h_{\frac{g_i}{d}}} = \frac{n}{d} - 1, \forall i \in \{1, 2, 3\}$ hence $I_{h_{g_i}} = n - d, \forall i \in \{1, 2, 3\}$.
- If $\frac{n}{d}$ is even then $I_{h_{\frac{g}{d}}} = \frac{n}{d} - 1$ and $I_{h_{\frac{g_3}{d}}} = \frac{n}{d} - 2$ therefore $I_{h_g} = n - d$ and $I_{h_{g_3}} = n - 2d$ by multiplying by d .

In addition to that, we have:

$$I_{n,k} = \frac{k.n}{2} - \frac{1}{2} \left(\sum_{i=1}^3 \gcd(n, g_i) \right) = \frac{3.n}{2} - \frac{1}{2} (d + d + 2d) = \frac{3n - 4d}{2} = d.I_{\frac{n}{d},k}$$

Or

$$I_{n,k} = \frac{k.n}{2} - \frac{1}{2} \left(\sum_{i=1}^3 \gcd(n, g_i) \right) = \frac{3.n}{2} - \frac{1}{2} (d + d + d) = \frac{3n - 3d}{2} = d.I_{\frac{n}{d},k}.$$

□

We denote by V the final matrix of synchronization for the co prime generators while V_d represent the matrix of synchronizations for generators with $\gcd \geq 2$ where $d = \gcd(g, g_3)$.

Lemma 6.17 *The matrix V_d is obtained by multiplying all the $v_{i,j}$'s of the matrix V by d and by concatenating the matrix obtained d times and we denote:*

$$V_d = d.V^{|d|}.$$

Proof. Let the triplet $G' = (g', g', g'_3) \pmod{n'}$ such that $n' = 2g' + g'_3$ and the triplet $G = (g, g, g_3) \pmod{n}$ such that $n = 2g + g_3$ where $g' = g.d$; $g'_3 = g_3.d$, $n' = n.d$ and $d = \gcd(g', n) = \gcd(g'_3, n)$.

Hence the orbits of g in $\mathbb{Z}/n\mathbb{Z}$ are:

$$\begin{aligned} g : 0 &\rightarrow g \rightarrow 2g \rightarrow \cdots \rightarrow (n-1)g \rightarrow 0 \\ g : 0 &\rightarrow g \rightarrow 2g \rightarrow \cdots \rightarrow (n-1)g \rightarrow 0 \\ g_3 : 0 &\rightarrow g_3 \rightarrow 2g_3 \rightarrow \cdots \rightarrow (n-1)g_3 \rightarrow 0 \end{aligned}$$

By multiplying by d we get:

$$\begin{aligned} d.g : 0 &\rightarrow g.d \rightarrow 2g.d \rightarrow \cdots \rightarrow (n-1)g.d \rightarrow 0 \\ d.g : 0 &\rightarrow g.d \rightarrow 2g.d \rightarrow \cdots \rightarrow (n-1)g.d \rightarrow 0 \\ d.g_3 : 0 &\rightarrow g_3.d \rightarrow 2g_3.d \rightarrow \cdots \rightarrow (n-1)g_3.d \rightarrow 0 \end{aligned}$$

\Leftrightarrow

$$\begin{aligned} g' : 0 &\rightarrow g' \rightarrow 2g' \rightarrow \cdots \rightarrow (n-1)g' \rightarrow 0 \\ g' : 0 &\rightarrow g' \rightarrow 2g' \rightarrow \cdots \rightarrow (n-1)g' \rightarrow 0 \\ g'_3 : 0 &\rightarrow g'_3 \rightarrow 2g'_3 \rightarrow \cdots \rightarrow (n-1)g'_3 \rightarrow 0 \end{aligned}$$

and since we multiplied by d so we are now working in $\mathbb{Z}/n'\mathbb{Z}$ so each row must be of length n' which means to repeat each row d times to reach the length n' . Hence we have: $V_d = d.V^{|d|}$. \square

Example

Take $(3, 3, 6) \pmod{12}$, the orbits of $g = 3$ and $g = 6$ in $\mathbb{Z}/12\mathbb{Z}$ are given as follow:

$$\begin{aligned} g_1 = 3 : 0 &\xrightarrow{*} 3 \xrightarrow{*} 6 \xrightarrow{*} 9 \xrightarrow{1} 0 \xrightarrow{*} 3 \xrightarrow{*} 6 \xrightarrow{*} 9 \xrightarrow{1} 0 \xrightarrow{*} 3 \xrightarrow{*} 6 \xrightarrow{*} 9 \xrightarrow{1} 0 \\ g_2 = 3 : 0 &\xrightarrow{*} 3 \xrightarrow{*} 6 \xrightarrow{*} 9 \xrightarrow{1} 0 \xrightarrow{*} 3 \xrightarrow{*} 6 \xrightarrow{*} 9 \xrightarrow{1} 0 \xrightarrow{*} 3 \xrightarrow{*} 6 \xrightarrow{*} 9 \xrightarrow{1} 0 \\ g_3 = 6 : 0 &\xrightarrow{*} 6 \xrightarrow{3} 0 \xrightarrow{*} 6 \xrightarrow{3} 0 \xrightarrow{*} 6 \xrightarrow{3} 0 \xrightarrow{*} 6 \xrightarrow{3} 0 \xrightarrow{*} 6 \xrightarrow{3} 0 \xrightarrow{*} 6 \xrightarrow{3} 0 \xrightarrow{*} 6 \xrightarrow{3} 0 \end{aligned}$$

We have: $I_{n,k} = \frac{3.12}{2} - \frac{1}{2}(3 + 3 + 6) = 12$; $I_{h_3} = 9$ and $I_{h_6} = 6$.

While $(1, 1, 2) \bmod 4$ has $I_{n,k} = \frac{3 \cdot 4 - 4}{2} = 4$; $I_{h_1} = 3$ and $I_{h_2} = 2$ where the orbits are:

$$\begin{aligned} g_1 &= 1 : 0 \xrightarrow{*} 1 \xrightarrow{*} 2 \xrightarrow{*} 3 \xrightarrow{1} 0 \\ g_2 &= 1 : 0 \xrightarrow{*} 1 \xrightarrow{*} 2 \xrightarrow{*} 3 \xrightarrow{1} 0 \\ g_3 &= 2 : 0 \xrightarrow{*} 2 \xrightarrow{3} 0 \xrightarrow{*} 2 \xrightarrow{3} 0 \end{aligned}$$

And hence we can write:

$$\begin{aligned} g_1 &= 3 : 3 \xrightarrow{*} 6 \xrightarrow{*} 9 \xrightarrow{1} 0 \xrightarrow{*} 3 \xrightarrow{*} 6 \xrightarrow{*} 9 \xrightarrow{1} 0 \xrightarrow{*} 3 \xrightarrow{*} 6 \xrightarrow{*} 9 \xrightarrow{1} 0 \xrightarrow{*} 3 \\ g_2 &= 3 : 9 \xrightarrow{1} 0 \xrightarrow{*} 3 \xrightarrow{*} 6 \xrightarrow{*} 9 \xrightarrow{1} 0 \xrightarrow{*} 3 \xrightarrow{*} 6 \xrightarrow{*} 9 \xrightarrow{1} 0 \xrightarrow{*} 3 \xrightarrow{*} 6 \xrightarrow{*} 9 \\ g_3 &= 6 : 0 \xrightarrow{*} 6 \xrightarrow{3} 0 \xrightarrow{*} 6 \xrightarrow{3} 0 \xrightarrow{*} 6 \xrightarrow{3} 0 \xrightarrow{*} 6 \xrightarrow{3} 0 \xrightarrow{*} 6 \xrightarrow{3} 0 \xrightarrow{*} 6 \xrightarrow{3} 0 \end{aligned}$$

which is equivalent to

$$= 3 \cdot \begin{pmatrix} 1 \xrightarrow{*} 2 \xrightarrow{*} 3 \xrightarrow{1} 0 \xrightarrow{*} 1 \\ 3 \xrightarrow{2} 0 \xrightarrow{*} 1 \xrightarrow{*} 2 \xrightarrow{*} 3 \\ 0 \xrightarrow{*} 2 \xrightarrow{3} 0 \xrightarrow{*} 2 \xrightarrow{3} 0 \end{pmatrix}^{|3}$$

6.6 Relation between $I_{n,k}$ and $I_{h_{g_i}}$

To summarize, whenever we have two equal generators, we are able to calculate the values of $I_{n,k}$ and $I_{h_{g_i}}$, $\forall i \in \{1, 2, 3\}$ where $\frac{n}{d}$ is either an even or odd number with $d = \gcd(n, g_i) \forall i$. To unify the work, we always let the row that starts with 0 corresponds to the generator g_3 that is distinct from the two others.

Lemma 6.18 *For thr triplet $G(g_1, g_2, g_3)$ with $n = g_1 + g_2 + g_3$ and $d = \gcd(n, g_i) \forall i \in \{1, 2, 3\}$,*

1. *Case A: If $\frac{n}{d}$ is odd, we get:*

$$I_{n,k} = 0 + \frac{I_{h_{g_i}}}{2} + I_{h_{g_i}} = 0 + \left(\frac{n-d}{2}\right) + (n-d) = \frac{3n-3d}{2},$$

and $S = \begin{pmatrix} \frac{I_{h_{g_i}}}{2} \\ I_{h_{g_i}} \\ 0 \end{pmatrix}$ represents its seed.

2. Case B: If $\frac{n}{d}$ is even, and we get:

$$I_{n,k} = 0 + \frac{I_{hg_3}}{2} + I_{hg} = 0 + \left(\frac{n-2d}{2}\right) + (n-d) = \frac{3n-4d}{2},$$

$$\text{and } S = \begin{pmatrix} \frac{I_{hg_3}}{2} \\ I_{hg} \\ 0 \end{pmatrix} \text{ its seed.}$$

Proof. Case A :

We define the three sets A, B, C to be the sets of decreasing positions for each generator starting respectively by $s_i \forall i \in \{1, 2, 3\}$.

$$A = \{i\alpha + x; i = 1 \dots \frac{g}{d}\}$$

$$B = \{i\alpha + y; i = 1 \dots \frac{g}{d}\}$$

$$C = \{i\beta; i = 1 \dots \frac{g_3}{d}\}$$

where: $\alpha g \equiv -d \pmod n$; $\beta g_3 \equiv -d \pmod n$ and $xg \equiv d \pmod n$ and $2yg \equiv 1 \pmod n$ since

$$gy \equiv -\frac{n-d}{2} \pmod n$$

$$2yg \equiv -(n-d) \pmod n$$

$$2yg \equiv d \pmod n$$

We need to prove that the sets A, B, C are pairwise disjoint, hence we can't get any two decreasing positions on the same column.

$$\underline{A \cap B = \emptyset}$$

Let $a \in A, b \in B$ we have $a := i\alpha + x$ and $b := j\alpha + y$ for some $i \in \{1 \dots \frac{g}{d}\}$ and $j \in \{1 \dots \frac{g}{d}\}$

By contradiction, suppose $a - b \equiv 0$ then:

$$(i-j)\alpha + x - y \equiv 0 \pmod n$$

$$2(i-j)g\alpha \equiv 2yg - 2gx \pmod n$$

$$2d(j-i) \equiv d \pmod n$$

This leads to a contradiction since we know that: $i - j \in \{\frac{1-g}{d}, \dots, \frac{g-1}{d}\}$, hence $2(i - j) = ln + d; l \in \mathbb{Z}$. If

- $l = 0; 2d(i - j) = d$ contradiction in \mathbb{Z} .
- $l > 0; 2d(i - j) > n + d$ contardiction.
- $l < 0; 2d(i - j) < -n + d$ contardiction.

therefore $A \cap B = \emptyset$.

$A \cap C = \emptyset$

We let $a \in A$ and $c \in C$; $a := i\alpha + x$ and $c := k\beta$, for some $i \in \{1, \dots, \frac{g}{d}\}$ and $k \in \{1, \dots, \frac{g_3}{d}\}$.

By contradiction, suppose $a - c \equiv 0 \pmod n$ then:

$$\begin{aligned} i\alpha + x - k\beta &\equiv 0 \pmod n \\ i\alpha g + xg - k\beta(\frac{n - g_3}{2}) &\equiv 0 \pmod n \\ -id + d - \frac{k\beta n}{2} - \frac{kd}{2} &\equiv 0 \pmod n \\ -2id + 2d - k\beta n - kd &\equiv 0 \pmod n \\ -2id + 2d - kd &\equiv 0 \pmod n \\ 2d(1 - i) - kd &\equiv 0 \pmod n \end{aligned}$$

this result leads to a contradiction since:

$$\begin{aligned} 1 \leq i \leq \frac{g}{d} \\ d - g \leq d(1 - i) \leq 0 \quad \text{and} \quad -g_3 \leq -dk \leq -d \\ 2d - 2g \leq 2d(1 - i) \leq 0 \end{aligned}$$

By adding both inequalities, we get: $-n < 2d - n \leq 2(1 - i) - k \leq -d < 0$, this is impossible since by hypothesis $2(1 - i) - k \equiv 0 \pmod n$, therefore the two sets A and C are disjoint.

$B \cap C = \emptyset$

We let $b \in B$ and $c \in C$; $b := i\alpha + y$ and $c := k\beta$, for some $i = 1 \dots \frac{g}{d}$ and $k = 1 \dots \frac{g_3}{d}$.

By contradiction, suppose $b - c \equiv 0 \pmod n$ then:

$$\begin{aligned}
i\alpha + y - k\beta &\equiv 0 \pmod n \\
i\alpha g + yg - k\beta\left(\frac{n - g_3}{2}\right) &\equiv 0 \pmod n \\
2i\alpha g + 2yg - k\beta n + k\beta g_3 &\equiv 0 \pmod n \\
-2id + d - k\beta n - kd &\equiv 0 \pmod n \\
-2id + d(1 - k) &\equiv 0 \pmod n
\end{aligned}$$

this result leads to a contradiction since:

$$\begin{aligned}
1 \leq i \leq \frac{g}{d} \\
-2g \leq -2id \leq -2d \quad \text{and} \quad d - g_3 \leq d(1 - k) \leq 0
\end{aligned}$$

By adding both inequalities, we get: $-n < d - n \leq -2id + d(1 - k) \leq -2d < 0$, this is impossible since by hypothesis $-2id + d(1 - k) \equiv 0 \pmod n$, therefore B and C are disjoint.

Case B :

When $\frac{n}{d}$ is an even number, we have: $\gcd(g, n) = d$ and $\gcd(g_3, n) = 2d$.

We define the sets A, B, C the sets of decreasing positions for each generator respectively starting by $s_i \forall i$.

$$\begin{aligned}
A &= \{i\alpha + x; i = 1, \dots, \frac{g}{d}\} \\
B &= \{i\alpha + y; i = 1, \dots, \frac{g}{d}\} \\
C &= C_0 \cup C_1
\end{aligned}$$

where $C_0 = \{i\beta\}$ and $C_1 = \{i\beta + j\frac{n}{2}\}; i = 1, \dots, \frac{g_3}{2d}\}$, $\alpha.g \equiv -d \pmod n$, $\beta.g_3 \equiv -2d \pmod n$ and $x.g \equiv d \pmod n$, $y.g \equiv d + m \pmod n$ where $n = 2m$, since it is an even number and we know that $y.g \equiv -s_1 \pmod n$ then:

$$\begin{aligned}
y.g &\equiv -s_1 \pmod n \\
y.g &\equiv -\frac{n - 2d}{2} \pmod n \\
y.g &\equiv -\frac{2m - 2d}{2} \pmod n \\
y.g &\equiv d - m \pmod n \\
y.g &\equiv d + m \pmod n
\end{aligned}$$

Now we prove that three sets are pairwise distinct starting with A and B:

$$\underline{A \cap B = \emptyset}$$

Let $a \in A, b \in B$, they are defined as: $a := i\alpha + x$ and $b := j\alpha + y$ where $i, j \in \{1, \dots, \frac{g}{d}\}$

By contradiction, suppose $a - b \equiv 0 \pmod n$ then:

$$\begin{aligned} i\alpha + x - j\alpha - y &\equiv 0 \pmod n \\ i\alpha.g + x.g - j\alpha.g - y.g &\equiv 0 \pmod n \\ -id + d + jd - d - m &\equiv 0 \pmod n \\ d(j - i) - m &\equiv 0 \pmod n \end{aligned}$$

This congruence leads to a contradiction since:

$$\begin{aligned} d - g &\leq d(j - i) \leq g - d \\ d - g - m &\leq d(j - i) - m \leq g - d - m \\ -n < d - g - m &\leq d(j - i) - m \leq g - d - m < 0 \end{aligned}$$

knowing that $m = \frac{n}{2} > g$. Hence $A \cap B = \emptyset$.

$$\underline{A \cap C = \emptyset}$$

In this part we divide our work into two parts since C is a union of two sets and we prove that A is disjoint with both subsets of C . We let $a \in A, c_j \in C_j \forall j \in \{0, 1\}$, they are defined as: $a := i\alpha + x$ and $c_0 := j\beta$, $c_1 = j\beta + \frac{n}{2}$ where $i \in \{1, \dots, \frac{g}{d}\}, j \in \{1, \dots, \frac{g_3}{2d}\}$

By contradiction, suppose $a - c_j \equiv 0 \pmod n$ then:

$$\begin{aligned} a - c_0 &\equiv 0 \pmod n \\ i\alpha + x - j\beta &\equiv 0 \pmod n \\ i\alpha.g + x.g - j\beta.\frac{n - g_3}{2} &\equiv 0 \pmod n \\ 2i\alpha.g + 2x.g - j\beta.n + j\beta.g_3 &\equiv 0 \pmod n \\ -2id + 2d - 2jd &\equiv 0 \pmod n \\ 2d - 2d(i + j) &\equiv 0 \pmod n \end{aligned}$$

For c_1 we get:

$$\begin{aligned}
a - c_1 &\equiv 0 \pmod{n} \\
i\alpha + x - j\beta - \frac{n}{2} &\equiv 0 \pmod{n} \\
i\alpha.g + x.g - j\beta.\frac{n - g_3}{2} - \frac{n}{2}.g &\equiv 0 \pmod{n} \\
2i\alpha.g + 2x.g - j\beta.n + j\beta.g_3 - ng &\equiv 0 \pmod{n} \\
-2id + 2d - 2jd &\equiv 0 \pmod{n} \\
2d - 2d(i + j) &\equiv 0 \pmod{n}
\end{aligned}$$

These two congruences lead to a contradiction since:

$$\begin{aligned}
2 &\leq d(i + j) \leq g + \frac{g_3}{2} \\
4d &\leq 2d(i + j) \leq n \\
-n &< 2d - n \leq 2d - 2d(i + j) \leq -2d < 0
\end{aligned}$$

Hence $A \cap C = \emptyset$.

$B \cap C = \emptyset$

Finally for B and C , we will repeat the same work by letting $b \in B$ and $c_j \in C_j \forall j \in \{0, 1\}$, they are defined as: $b := i\alpha + y$ and $c_0 := j\beta, c_1 = j\beta + \frac{n}{2}$ where $i \in \{1, \dots, \frac{g}{d}\}, j \in \{1, \dots, \frac{g_3}{2d}\}$.

By contradiction, suppose $b - c_j \equiv 0 \pmod{n}$ then:

$$\begin{aligned}
b - c_0 &\equiv 0 \pmod{n} \\
i\alpha + y - j\beta &\equiv 0 \pmod{n} \\
i\alpha.g + y.g - j\beta.\frac{n - g_3}{2} &\equiv 0 \pmod{n} \\
2i\alpha.g + 2y.g - j\beta.n + j\beta.g_3 &\equiv 0 \pmod{n} \\
-2id + 2(d + m) - 2jd &\equiv 0 \pmod{n} \\
-2id + 2d + 2m - 2jd &\equiv 0 \pmod{n} \\
2d - 2d(i + j) &\equiv 0 \pmod{n}
\end{aligned}$$

For c_1 we get:

$$\begin{aligned}
a - c_1 &\equiv 0 \pmod{n} \\
i\alpha + y - j\beta - \frac{n}{2} &\equiv 0 \pmod{n} \\
i\alpha.g + y.g - j\beta.\frac{n - g_3}{2} - \frac{n}{2}.g &\equiv 0 \pmod{n} \\
2i\alpha.g + 2y.g - j\beta.n + j\beta.g_3 - ng &\equiv 0 \pmod{n} \\
-2id + 2(d + m) - 2jd &\equiv 0 \pmod{n} \\
-2id + 2d + 2m - 2jd &\equiv 0 \pmod{n} \\
2d - 2d(i + j) &\equiv 0 \pmod{n}
\end{aligned}$$

These two equivalences are the same as the previous case hence we get a contradiction too and the sets B and C are disjoint. \square

6.7 Equal generators

After showing and studying the synchronization for the case where two of the three generators are equal, the trivial questions that arise immediately to the mind are the following: What happened if the three generators are equal? Are we also able to synchronize them?

In fact, the answer is positive and to prove it, we let $g_1 = g_2 = g_3 = g$ and we can conclude the following:

Lemma 6.19 *Let $G = (g, g, g)$ be a triplet formed of the same generator g . Orbits in $\mathbb{Z}/n\mathbb{Z}$ where $n = 3g$ are always of the form:*

$$V = g \begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{pmatrix}^{|g|}$$

and the word formed is always of the form $(321)^g$.

Proof. For equal generators, we have $G = (g, g, g) \pmod{3g}$. Each row of V has exactly g decreasing positions, $2g$ increasing positions and is of length $3g$. By calculating the orbits of g in $\mathbb{Z}/3\mathbb{Z}$, we remark that each row is written as $(0 \ g \ 2g)^{|g|}$. Factoring by g , we get: $g \ (012)^g$. To accomplish the synchronization, we do a

$k - 1$ circular permutation $\forall k \in \{1, 2, 3\}$, hence we get: $g \begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{pmatrix}^{lg}$ and the word formed is: $(321)^g$. \square

Proposition 6.20 *For the equal generators we have:*

1. The seed is given by: $S = \begin{pmatrix} 0 \\ g \\ 2g \end{pmatrix}$.
2. The vertical invariant is: $I_{n,k} = 3g$.
3. The horizontal invariant is: $I_{h_{g_i}} = 2g \forall i \in \{1, 2, 3\}$.

Proof. Since the synchronization matrix is of the form $V = g \begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{pmatrix}^{lg}$,

then the seed in this case has always the form: $S = \begin{pmatrix} 0 \\ g \\ 2g \end{pmatrix}$. Using the Lemma

6.11, the value of the horizontal invariant is calculated as follows and gives: $n - d = 3g - g = 2g$. Since $I_{n,k}$ is the vertical sum of the elements of each column so in particular we can get the value from the seed, hence $I_{n,k} = 0 + g + 2g = 3g$. Or simply by computing:

$$I_{n,k} = \frac{k \cdot n}{2} - \frac{1}{2} \left(\sum_{i=1}^3 \gcd(n, g_i) \right) = \frac{9g}{2} - \frac{1}{2}(g + g + g) = \frac{6g}{2} = 3g.$$

\square

6.8 Distinct generators

The last case to check for the synchronization of three generators is when the three generators are pairwise distinct. In fact, this case can be covered by using the **Fraenkel's** conjecture see [25, 51] for $k = 3$ that states the uniqueness of synchronization for pairwise distinct generators when $g_i = 2^{i-1} \forall i \in \{1, \dots, k\}$ that was proved for $3 \leq k \leq 7$. Therefore there exists a **unique** balanced word on $3 \leq k$ letters, with pairwise distinct generators obtained from the triplet $(1, 2, 4) \bmod 7$ which is the unique solution for this case. Example 78 shows that:

1. $I_{n,k} = \frac{3n-3}{2} = \frac{21-3}{2} = 9$ since n is odd.
2. $I_{h_1} = I_{h_2} = I_{h_4} = n - 1 = 6$ since all the generators are co prime with n .
3. Finally, the word formed is $w = 2313233$.

REMARK: All the multiples of $G = (1, 2, 4)$ are also synchronized since we can write: $G' = dG^{|d|}$, where d is the greatest common divisor between the three pairwise distinct generators.

6.8.1 Fraenkel's seed

For the case of pairwise distinct generators, we stated that the unique solution corresponds to the Fraenkel's case. We give in this part and in a more general way the seed for the generators of the form $2^i \forall i \in \{0, \dots, m\}$.

Theorem 6.21 *The seed for the sequence of generators of the form $2^i \forall i \in \{0, \dots, m\}$ is: $S = \begin{pmatrix} \frac{I_{n,k}}{k} \\ \vdots \\ \frac{I_{n,k}}{k} \end{pmatrix}$ where k is the number of elements in this sequence.*

Proof. Let the sequence of generators of the form $2^i \forall i \in \{0, \dots, m\}$ and $n = 2^{m+1} - 1$ their sum.

We define the sets A and B of decreasing position for the two generators : $g_l = 2^l$ and $g_s = 2^s$ where $l < s \leq m$ starting with the seed s_l, s_s and respecting all the conditions given in the previous sections by the following:

We know that: $I_{n,k} = \frac{k \cdot n}{2} - \frac{1}{2}(k) = \frac{k(n-1)}{2}$ hence:

$$s_i = \frac{I_{n,k}}{k} = \frac{n-1}{2} \forall i \in \{0, \dots, m\} \text{ where } \frac{n-1}{2} = 2^m - 1.$$

$$A = \{i\alpha + x; i = 1 \dots 2^l\}$$

$$B = \{j\beta + y; j = 1 \dots 2^s\}$$

where: $\alpha 2^l \equiv -1 \pmod{n}$; $\beta 2^s \equiv -1 \pmod{n}$ and $x \cdot 2^l \equiv 1 - 2^m \pmod{n}$ and $y \cdot 2^s \equiv 1 - 2^m \pmod{n}$ since $x \cdot g$ must be equivalent to $-s = -(2^m - 1) = 1 - 2^m$.

We need to prove that the sets A, B are disjoint, hence we can't get any two decreasing positions on the same column.

$$\underline{A \cap B = \emptyset}$$

Let $a \in A, b \in B$ we have $a := i\alpha + x$ and $b := j\beta + y$ for some $i \in \{1 \dots 2^l\}$ and

$$j \in \{1 \dots 2^s\}$$

By contradiction, suppose $a - b \equiv 0 \pmod n$ then:

$$\begin{aligned} i\alpha + x - j\beta - y &\equiv 0 \pmod n \\ 2^s i\alpha + 2^s x - 2^s j\beta - 2^s y &\equiv 0 \pmod n \\ -2^{s-l}i + 2^{s-l}(1 - 2^m) + j - 1 + 2^m &\equiv 0 \pmod n \\ -2^{s-l}i + 2^{s-l} - 2^{m+s-l} + j - 1 + 2^m &\equiv 0 \pmod n \\ 2^{s-l}(1 - i) - 2^{m+s-l} + j - 1 + 2^m &\equiv 0 \pmod n \end{aligned}$$

where:

$$1 \leq i \leq 2^l \\ 2^{s-l} - 2^s \leq 2^{s-l}(1 - i) \leq 0 \quad \text{and} \quad -0 \leq j - 1 \leq 2^s - 1$$

By adding both inequalities with the constant term in $a - b$ we get:

$$2^m - 2^{m+s-l} - 2^s + 2^{s-l} \leq a - b \leq 2^s - 1 + 2^m - 2^{m+s-l}.$$

Our aim is to reach a contradiction by bounding the two terms A and B of this inequality by two consecutive multiples of n and hence the term $a - b$ who is bounded by $A \leq a - b \leq B$ can't be null.

The power $m + s - l$ is the highest power in the two terms of A and B, $m + 1$ is the power found in n . We let $p = (m + s - l) - (m + 1) = s - l - 1$ and $k = 2^p$, we will prove that: $-k.n < A \leq a - b \leq B < (-k + 1).n$

First, $-k.n = -2^{s-l-1}.(2^{m+1} - 1) = -2^{m+s-l} + 2^{s-l-1}$, then $-k.n < A$ will give us the following:

$$\begin{aligned} -2^{m+s-l} + 2^{s-l-1} &< 2^m - 2^{m+s-l} - 2^s + 2^{s-l} \\ 2^{s-l-1} - 2^m + 2^s - 2^{s-l} &< 0 \\ 2^{s-l}(2^{-1} - 1) - 2^m + 2^s &< 0 \\ -2^{s-l-1} - 2^m + 2^s &< 0 \end{aligned}$$

Which is always true since $l < s \leq m$.

Now, we prove that $B < (-k + 1).n$ by calculating at the beginning the value of $(-k + 1).n$, we get the following :

$$(-k + 1).n = (-2^{s-l-1} + 1)(2^{m+1} - 1) = -2^{s-l+m} + 2^{s-l-1} + 2^{m+1} - 1 \text{ then,}$$

$$\begin{aligned} 2^{s-l+m} + 2^{s-l-1} + 2^{m+1} - 1 &> 2^s - 1 + 2^m - 2^{m+s-l} \\ 2^{s-l-1} + 2^{m+1} - 2^s - 2^m &> 0 \\ 2^{s-l-1} - 2^s + 2^m &> 0 \end{aligned}$$

which is always true since:

$$2^l < 2^s \leq 2^m \\ -2^m \leq -2^s < -2^l \quad \text{and} \quad 2^0 \leq 2^{s-l-1} < 2^m$$

by adding these two inequalities we get: $1 - 2^m \leq 2^{s-l-1} - 2^s < 2^m - 2^l < 2^m$ confirming the positivity of $2^{s-l-1} - 2^s + 2^m$.

All this work leads to a contradiction since the term $a - b$ can't be equivalent to 0. \square

6.9 Relation between 2 and 3 generators

Till now we are able to synchronize two and three generators but the following question derives: Are we able to find a relation that links these two synchronizations? In other words, are we able to get the synchronizations of two generators if the the synchronization of three generators is given? We let $G = (g, g, g_3)$ be a triplet of three generators and we study the two cases depending on the parity of n where $n = 2g + g_3$.

1. If $\frac{n}{d}$ is **odd**:

In this case, the seed is of the form: $S = \begin{pmatrix} \frac{n-d}{2} \\ n-d \\ 0 \end{pmatrix}$ where the row that starts with 0 refers to g_3 . Since we know that any two generators of the form $G' = (2g, g_3)$ can be synchronized with seed $S' = \begin{pmatrix} n-d \\ 0 \end{pmatrix}$ and where the row starting with 0 belongs to g_3 , then the synchronization of G' is obtained by adding the value $\frac{n+d}{2}$ to the sum of the first two synchronized rows of G .

$$\begin{aligned} n-d + \frac{n-d}{2} + x &= n-d \pmod{n} \\ \frac{n-d}{2} + x &= n \pmod{n} \\ x &= \frac{n+d}{2} \pmod{n}. \end{aligned}$$

Example 80 *Passing from the synchronization of (3, 3, 5) to the one of (6, 5): Let (3, 3, 5) mod 11 we can write its configuration like the following:*

$$V = \begin{array}{c|c|c|c|c|c|c|c|c|c|c} 5 & 8 & 0 & 3 & 6 & 9 & 1 & 4 & 7 & 10 & 2 \\ 10 & 2 & 5 & 8 & 0 & 3 & 6 & 9 & 1 & 4 & 7 \\ 0 & 5 & 10 & 4 & 9 & 3 & 8 & 2 & 7 & 1 & 6 \end{array}$$

If we add the first two rows we get:

$$4 \mid 10 \mid 5 \mid 0 \mid 6 \mid 1 \mid 7 \mid 2 \mid 8 \mid 3 \mid 9$$

And if we write the configuration of $(6, 5) \pmod{11}$ we get:

$$V' = \begin{array}{c} 10 \mid 5 \mid 0 \mid 6 \mid 1 \mid 7 \mid 2 \mid 8 \mid 3 \mid 9 \mid 4 \\ 0 \mid 5 \mid 10 \mid 4 \mid 9 \mid 3 \mid 8 \mid 2 \mid 7 \mid 1 \mid 6 \end{array}$$

We can realize that by adding $+6$ to the sum of the two first rows of V , we get immediately the first row of V' since: $\frac{n+1}{2} = \frac{12}{2} = 6$ and in this case $d = 1$.

2 If $\frac{n}{d}$ is even:

In this case, similarly we have: $S = \begin{pmatrix} \frac{n-2d}{2} \\ n-d \\ 0 \end{pmatrix}$ is a seed where 0 belongs to the row of g_3 . Since $\frac{n}{d}$ is even, then $\gcd(g_3, n) = 2d$ and $\gcd(g, n) = d$; hence the seed for $G' = (2g, g_3)$ will be: $S' = \begin{pmatrix} 0 \\ n-2d \end{pmatrix}$. The value to be added to the sum of the two rows of equal generator is $\frac{n}{2}$ since we have:

$$\begin{aligned} n-d + \frac{n-2d}{2} + x &= n-2d \pmod{n} \\ \frac{n-2d}{2} + x &= -d \pmod{n} \\ x &= \frac{n}{2} \pmod{n}. \end{aligned}$$

Example 81 Starting with the three generators $(3, 3, 4) \pmod{10}$ we can easily write its configuration like the following:

$$V = \begin{array}{c} 4 \mid 7 \mid 0 \mid 3 \mid 6 \mid 9 \mid 2 \mid 5 \mid 8 \mid 1 \\ 9 \mid 2 \mid 5 \mid 8 \mid 1 \mid 4 \mid 7 \mid 0 \mid 3 \mid 6 \\ 0 \mid 4 \mid 8 \mid 2 \mid 6 \mid 0 \mid 4 \mid 8 \mid 2 \mid 6 \end{array}$$

If we add the first two rows we get:

$$3 \mid 9 \mid 5 \mid 1 \mid 7 \mid 3 \mid 9 \mid 5 \mid 1 \mid 7$$

And if we write the configuration of $(6, 4) \pmod{10}$ we get:

$$V' = \begin{array}{c} 9 \mid 5 \mid 1 \mid 7 \mid 3 \mid 9 \mid 5 \mid 1 \mid 7 \mid 3 \\ 0 \mid 4 \mid 8 \mid 2 \mid 6 \mid 0 \mid 4 \mid 8 \mid 2 \mid 6 \end{array}$$

We can realize that by adding $+6$ to the sum of the first two rows of V , we get immediately the first row of V' since: $\frac{n+2}{2} = \frac{12}{2} = 6$.

Open questions Finally we can say that any three generators can be transformed to a couple of two generators that are certainly synchronized. We can resume the study by the following:

Let $(g_1, g_2, g_3) \pmod n$:

- If all the generators are pairwise distinct \longrightarrow The Fraenkel's case for $k = 3$ is the unique solution.
- If at least two generators are not distinct \longrightarrow Always can be synchronized.

Many questions can arise after this work:

- i Can we study the case of 4 generators and see when they can be synchronized?
- ii Using the second order balanced matrix, we are able to get some information about the behavior of Christoffel words. With the synchronized rows in this matrix, we can think of extending the work of synchronization to the k^{th} case?
- iii Can we apply this result in geometry by considering each generator as a direction vector of a discrete line in kD and see what may happen?
- iv Are we able to solve the Fraenkel's conjecture for $k > 7$ using also all the properties of the second order balance matrix?

Some perspectives:

The main purpose of this thesis is to find a way to synchronize several Christoffel words. In fact, it was not an easy and direct result or way to solve this problem. We started by introducing a new way to obtain the synchronization of three Christoffel words using the seed and we tried to introduce a new tool called the second order balance matrix that do a kind of classification of these balanced words. During the work many questions were proposed and several new ideas were treated. Now, after explaining the role of the seed and the two invariants, in this paragraph, we will show what is the link between this tool and the synchronization.

Figure 6.3 shows the three matrices $U_{\frac{1}{6}}, U_{\frac{2}{5}}$ and $U_{\frac{4}{3}}$. In fact, we start with the example of Fraenkel, using the three generators: $(1, 2, 4) \pmod n$ and as we can see in Example 78, this triplet is synchronized with $I_{n,k} = 9$ and with seed equals to:

$$S = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}.$$

We start to remark that in the second order balance matrix, we always have 4 or 3 particular rows depending on the parity of n that is the length of the word.

$$U_{\frac{1}{6}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 2 & 1 \\ 1 & 2 & 3 & 3 & 2 & 1 \\ 1 & 2 & 3 & 3 & 2 & 1 \\ 1 & 2 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}, \quad U_{\frac{2}{5}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$U_{\frac{3}{4}} = U_{\frac{3}{4}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 & 1 & 1 \\ 1 & 1 & 2 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Figure 6.3: The three matrices $U_{\frac{1}{6}}$, $U_{\frac{2}{5}}$ and $U_{\frac{3}{4}}$ used for the synchronization of the triplet $(1, 2, 4) \bmod 7$.

Those four rows are composed of the two synchronized rows, that are the rows containing only 1 and the two other rows are the rows that the sum of their elements gives a maximal value comparing to the other rows. We name $M_{\frac{a}{b}}$ the set of the number of the rows with a maximal value and $m_{\frac{a}{b}}$ is the set of the number of the synchronized rows. We have that:

1. $M_{\frac{1}{6}} = \{3, 4\}$ and $m_{\frac{1}{6}} = \{1, 6\}$;
2. $M_{\frac{2}{5}} = \{2, 5\}$ and $m_{\frac{2}{5}} = \{1, 3, 4, 6\}$;
3. $M_{\frac{3}{4}} = \{3, 4\}$ and $m_{\frac{3}{4}} = \{1, 2, 5, 6\}$.

Those numbers have a particular explication since if we use the seed where all its elements are equal to 3 and we calculate $1.x \equiv_7 3$, $2.x \equiv_7 3$ and $4.x \equiv_7 3$, we get the values 3, 5, 6 respectively. Those values are the elements of the sets $m_{\frac{1}{6}}$, $m_{\frac{2}{5}}$ and $m_{\frac{3}{4}}$ distinct from the rows 1 and $n - 1 = 6$. In addition to that we can also see that the six sets are related in such a way that the elements of $M_{\frac{1}{6}}$ appear in $m_{\frac{2}{5}}$ and $M_{\frac{3}{4}} = \{3, 4\}$. Furthermore, we can also realize that starting with the elements of the set $m_{\frac{3}{4}}$, if we double the values, distinct from 1 and $n - 1$, we get the elements of $m_{\frac{2}{5}}$. The same remark can be also given for the sets $m_{\frac{2}{5}}$ and $m_{\frac{1}{6}}$.

- $2.2 = 4$; $5.2 = 10 \equiv_7 3$, where $\{3, 4\} \subset m_{\frac{2}{5}}$.
- $4.2 = 8 \equiv_7 1$; $3.2 = 6$, where $\{1, 6\} = m_{\frac{1}{6}}$.

We will see that the same remarks can be given for three Christoffel words having two equal generators and one distinct. Let us try the triplet $(3, 3, 5) \bmod n$. We need in this case the two matrices: $U_{\frac{3}{8}}$ and $U_{\frac{5}{6}}$, where the synchronized form has

a seed equals to: $S = \begin{pmatrix} 5 \\ 10 \\ 0 \end{pmatrix}$. We recall that the last row with 0 corresponds to the distinct generator. By checking the two matrices in Figure 6.4, we can get the sets $M_{\frac{3}{8}}$, $M_{\frac{5}{6}}$, $m_{\frac{3}{8}}$ and $m_{\frac{5}{6}}$.

$$U_{\frac{3}{8}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 2 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 2 & 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 & 2 & 2 & 1 & 2 & 2 & 1 \\ 1 & 2 & 2 & 1 & 2 & 2 & 1 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 2 & 2 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 2 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \quad U_{\frac{5}{6}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 1 \\ 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 1 \\ 1 & 1 & 2 & 2 & 3 & 3 & 2 & 2 & 1 & 1 \\ 1 & 1 & 2 & 2 & 3 & 3 & 2 & 2 & 1 & 1 \\ 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 1 \\ 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Figure 6.4: The two matrices $U_{\frac{3}{8}}$ and $U_{\frac{5}{6}}$ used for the synchronization of the triplet $(3, 3, 5) \pmod{11}$.

The sets are exactly:

1. $M_{\frac{3}{8}} = \{5, 6\}$ and $m_{\frac{3}{8}} = \{4, 7\}$;
2. $M_{\frac{5}{6}} = \{5, 6\}$ and $m_{\frac{5}{6}} = \{1, 2, 9, 10\}$.

Again those values appear by computing: $3.x \equiv_{11} 5, 3.x \equiv_{11} 10, 5.x \equiv_{11} 10$ and $5.x \equiv_{11} 5$, and we get 9, 7, 2 and 1 respectively. By multiplying by 2 the elements of $m_{\frac{5}{6}}$ distinct from 1 and $n - 1 = 10$, we get the elements of $m_{\frac{3}{8}}$, since we have: $2.2 = 4$ and $9.2 = 18 \equiv_{11} 7$.

This work is not over yet, we are still looking for a general result to synchronize the Christoffel words using the second order balance matrix. This is a first step in our main goal and more small results are obtained but not yet proved, those results will be the point of interest for my further work after the thesis.

Chapter 7

Convexity and digital convex polyominoes

Before starting this chapter, I would like to thank my Italian colleagues: Paolo Dulio, Andrea Frosini and Simone Rinaldi, for their visit to the LAMA and our fruitful discussions that led to this result represented in Chapter 7. Together with my director Laurent Vuillon, had a paper about the reconstruction of the polyominoes, that was presented in the conference of WORDS 2017 in Montréal.

Studying families of words has started with Bernouilli at the beginning of 1771, then was followed by Markov, Thue and Morse. Those studies are applied in different domain of research, like number theory, group theory, probabilistic theory. We can't forget to mention the impact of all the results in the theoretical computer science field and in particular all the works dealing with automaton and formal languages. For a survey on the topic one can refer to Lothaire's book [33]. Finally a relation between combinatorics on words and discrete geometry has afforded many advantages to both areas and has led to interesting results. Indeed, such a connection realizes through the so called Freeman coding, introduced by Freeman in 1961, that allows to uniquely determine a 4 or 8 connected finite set of points in the discrete plane by means of its boundary word, i.e. a word over the alphabet of cardinal four $A = \{0, 1, \bar{0}, \bar{1}\}$. This code was the bridge between these two worlds. In addition, connected and without holes discrete sets or what is said polyominoes or their 8 connections are coded by words using the Freeman code. In this chapter, we introduce one of the main notions of convexity for polyominoes present in literature: the *Digital convexity* (further notions and related studies can be found in [10, 11, 16, 18, 19]) in order to present later on the first steps toward a reconstruction algorithm for those class of discrete sets. The first steps towards the

reconstruction of Digital convex (DC) sets play a prominent role in the framework of digital geometry providing a natural generalization to the concept of Euclidean convexity when we are dealing with polyominoes (who are the interior of a closed non intersecting grid path of \mathbb{Z}^2) [20]. My approach to the reconstruction problem relies on a recent result of Brlek, Lachaud, Provençal and Reutenauer (see [9]) that was a bridge between digitally convex notion and combinatorics on words. Indeed, a polyomino P is described by its border word $Bd(P)$. The border word $Bd(P)$ can be divided in 4 monotone paths. After computing the Lyndon factorization of each path, if each of the factors of the Lyndon factorizations is composed of Christoffel words then we have a digitally convex polyomino.

7.1 Introduction

It is well known that the digitalization process does not preserve connectedness so that it may happen that connected sets in the Euclidean plane can be transformed into two or more disconnected sets of points in the lattice \mathbb{Z}^2 . Let us label the pixels in a subset of a digital image, in such a way that any two pixels A and B have the same label if and only if they are connected, i.e if there exists a sequence of pixels $A = A_0, A_1 \dots A_n = B$ belonging to the same subset such that A_i is a neighbor of, or adjacent to, A_{i-1} for all $1 \leq i \leq n$. The points having the same label in a set S are called *connected component* of S . So, each set can be seen as the union of its connected components. We name a *hole* in S , a finite connected component of \bar{S} . A digital image has a convex subset S if the straight line segment joining any two pixels P and Q of the subset lies entirely in S . We introduce the convex hull of S denoted $conv(S)$ that is the smallest convex set containing S . It is not so easy to define digital convex sets since there is in general some difference between convex and digital convex. As an example, in the papers of Sklansky [48] and Minsky and Papert [48] digital convex sets may contain many connected components. In order to have exactly non connected component, Chaudhuri and Rosenfeld [12] impose implicitly that a digital convex set must be a polyomino. Recall that a *polyomino* P is a simply connected union of finite unit points that can be visualized as set of unit squares in a squared surface. In fact the two authors propose the notion of *DL-convexity* (where DL means digital line) and by definition a discrete region is *DL-convex* if, for any two squares belonging to it, there exists a digital line between them all of whose points belong to the region. Thus for Chaudhuri and Rosenfeld the region must be a polyomino. Debled-Rennesson, Rémy and Rouyer-Degli have worked on the arithmetic properties of discrete segments in order to detect the convexity of polyominoes (see [12]). In this section, we recall some basic notions about the *DC* that will be used to treat the problem we want to address:

How to insert a single point in a DC set in order to maintain the property of digital convexity? Similar problems fit in the framework of Discrete Tomography that is a part of the wide area of Computerized Tomography, and it studies the reconstruction of polyominoes.

To be able to understand the reconstruction part, we need to give an explicit definition to all the terminologies already used.

7.1.1 Discrete geometry and digital space

Let us consider the lattice \mathbb{Z}^2 , the canonical basis of the euclidean vector space \mathbb{R}^2 is $\{e_1, e_2\}$. We call a finite discrete set any subset of the discrete space \mathbb{Z}^2 which is the set of all the vectors having integer components. We define a path in \mathbb{Z}^2 from point A to point B as: the sequence of points $(p_i)_{1 \leq i \leq n}$ with $p_i \in \mathbb{Z}^2$ where we have $p_1 = A$, $p_n = B$ and such that p_i and p_{i+1} are neighbors with respect to some criteria.

We mention those two notions of neighboring:

- Let $p = (p_x, p_y)$ be a point of the discrete plane, the 4 neighborhood of a point p is given by:

$$\{(x, y) \in \mathbb{Z}^2 \mid |p_x - x| + |p_y - y| = 1\}.$$

In addition, a path in \mathbb{Z}^2 is said 4-connected if for all $1 \leq i \leq n - 1$, we have:

$$p_{i+1} - p_i \in \{\pm e_1, \pm e_2\}.$$

- Similarly, the 8 neighborhood of a point p is given by:

$$\{(x, y) \in \mathbb{Z}^2 \mid \max\{|p_x - x|, |p_y - y|\} = 1\}.$$

A path in \mathbb{Z}^2 is 8-connected if for all $1 \leq i \leq n - 1$, we have:

$$p_{i+1} - p_i \in \{\pm e_1, \pm e_2, \pm e_1 \pm e_2\}.$$

Figure 7.1 represents geometrically the 4 and 8 neighbors of a point in the \mathbb{Z}^2 grid.

We remark that in a 4-connected set all the elements, except the single point, have at least one 4 neighborhood but the reverse does not hold.

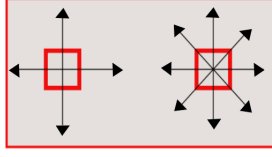


Figure 7.1: The left square shows the 4 possible directions of a point in a 4-connected path and the right square shows the 8 possible directions of a point in an 8-connected path.

Digital convexity During the last decades, many definitions were given of Digital convexity. It started with Minsky and Papert in 1969 [36] then Hubler. We show the evolution of the definitions and we conclude by the definition that will be used during this thesis.

1. **MP-convexity:** The definition given by Minsky and Papert states that: For x and y in $S \subset \mathbb{Z}^2$ and $z \in [x, y] \cap \mathbb{Z}^2 \Rightarrow z \in S$, where $[x, y]$ is the line segment joining x to y and defined by: $[x, y] = \{z \in \mathbb{R}^2 | z = \lambda x + (1 - \lambda)y, 0 \leq \lambda \leq 1\}$.
2. **H-convexity:** A set S is convex if it equals (the digitalization of) its convex hull, where the convex hull of S is defined by:

$$\text{conv}(S) = \left\{ \sum_{j=1}^n \lambda_j x_j \mid \sum_{j=1}^n \lambda_j = 1, \lambda_j \geq 0 \text{ and } x_j \in S \right\}.$$

S is **H-convex** if $S = \text{conv}(S) \cap \mathbb{Z}^2$.

3. The definition given by Hulber is the **DH-convexity**; S is said Digitally convex if for any $x, y \in S$, the line segment passing by x and y belongs to S . Hence, we remark that S is *MP-convex* if and only if S is *DH-convex*.
4. Finally, a set S is said *D-convex* if it is 8-connected and *H-convex* at the same time.

In this thesis, we use the last definition since we are working with digital convex polyomino. Figures 7.2 and 7.3 show two examples to visualize the difference between the different kinds of convexity.

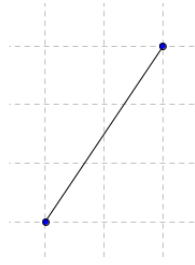


Figure 7.2: Example of an MP-convex and H-convex.

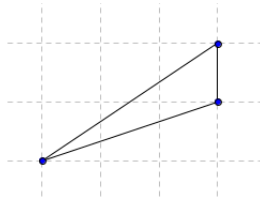


Figure 7.3: Example of an MP-convex but not H-convex.

Freeman chain code

The provided preliminary notions allow us to define the Freeman chain code that grants us to represent any path over a discrete plane \mathbb{Z}^2 as a word over an alphabet of four letters $A = \{0, 1, \bar{0}, \bar{1}\}$ [26] such that:

1. We denote 0 for a right horizontal step (\rightarrow).
2. We denote $\bar{0}$ for a left horizontal step (\leftarrow).
3. We denote 1 for an upside vertical step (\uparrow).
4. We denote $\bar{1}$ for an downward vertical step (\downarrow).

For a polyomino, the word obtained by applying the Freeman code to the perimeter of a polyomino P is called its boundary word, denoted $Bd(P)$. We are able to characterize a polyomino by its border word see Figure 7.4.

We remark that for the same polyomino, we can have several border words depending on the starting point of the word. So we are able to say that a polyomino is identified by an equivalent class of words that are all the cyclic shifts of the same one. We choose as representative of this equivalent class, the word that starts from

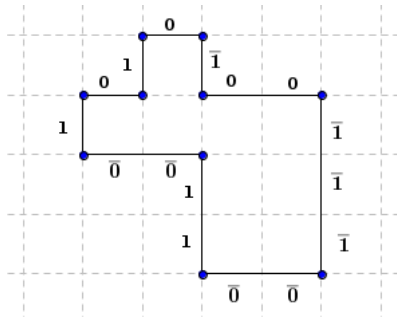


Figure 7.4: An example of a polyomino with border word: $w = Bd(P) = 010\bar{1}00\bar{1}\bar{1}\bar{1}00\bar{1}1\bar{1}00\bar{1}$.

the uppermost element. For example, in Figure 7.4, the border word can also be given by $w' = 00\bar{1}\bar{1}\bar{1}00\bar{1}1\bar{1}00\bar{1}010\bar{1}$. These two words encode the same polyomino and therefore we can say that they are two equivalent border words and one is obtained by a circular shift of the other. Hence we write: $w \equiv w'$.

Remark 7.1 Over the alphabet $A = \{0, 1, \bar{0}, \bar{1}\}$, a path with encoded word w is said to be close if A coincides with B , equivalently $A = B$, where A and B are the initial and ending points, i.e. we have that $p_0 = p_k$. We can observe that in this case, the numbers of 0s and $\bar{0}$ s are equal and the same holds for the number of 1s and $\bar{1}$ s. We name it in this case the AB path.

Example 82 In Figure 7.5, it is depicted a path in \mathbb{Z}^2 with its corresponding word encoded over the alphabet: $A = \{0, 1, \bar{0}, \bar{1}\}$ where $w = 1111\bar{0}\bar{0}\bar{1}\bar{1}\bar{1}000$.

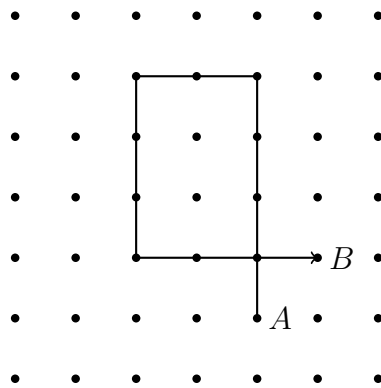


Figure 7.5: A path coded by the word: $w = 1111\bar{0}\bar{0}\bar{1}\bar{1}\bar{1}000$.

We remark that this path is not close since the initial and ending points, A and B respectively are not the same. We can also notice that for the border word, we have: 4 times the letter 1, while we have only 3 times the letter $\bar{1}$. Similarly, we have 3 times the letter 0, while we have only 2 times the letter $\bar{0}$.

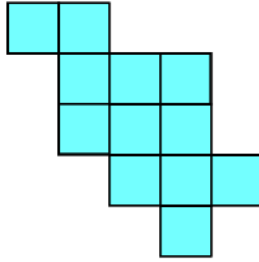


Figure 7.6: An example of a polyomino in the \mathbb{Z}^2 lattice.

Polyominoes

In this chapter, we deal with polyominoes that were introduced by Gardner in [27]. The word polyomino was firstly used by Golomb in [28], who produced a revised edition of Gardner's work and covered the enormous literature of polyominoes that has developed over the years. In this thesis, we choose to consider, as commonly accepted, only polyomino that is a 4-connected finite subset of \mathbb{Z}^2 having no holes and we define it as follows:

Definition 7.2 *A polyomino P is said digitally convex if the convex hull of P and \mathbb{Z}^2 are in P ; $\text{conv}(P) \cap \mathbb{Z}^2 \in P$.*

7.2 Theoretical results

Combinatorics on Words is a tool used in all the branches of mathematics. We will see in this section its contribution in discrete geometry. Using the Lyndon factorization and all the properties of Christoffel words stated in chapters 1 and 2, an interesting result about convexity is stated in [9].

Each of the four letters in the alphabet A used to code the border of a connected set, $Bd(S)$, provides a step along the border in one of the four different directions North, South, East and West, respectively as mentioned in the above section.

As we said before, each equivalence class of words that represents the border word of a set or a polyomino has a unique representative. In fact, to reach a standard coding of the border of $Bd(S)$, it can be noticed that a convex set touches the border of its minimal bounding rectangle in four bars, called *(N)orth*, *(S)outh*, *(E)ast* and *(W)est foot*. Moving clockwise on the border of the set, let us denote the ending corner of each foot by N , E , S and W according to the correspondent foot, as shown in Fig. 7.7. The word $Bd(S)$, that is the representative of the equivalence class, starts from W and runs clockwise along the border of S in a

closed path: $Bd(S)$ can be factorized into four non-void sub-paths WN , NE , ES and SW each using only two of the four steps in A to connect the related points; such a factorization is called *standard*.

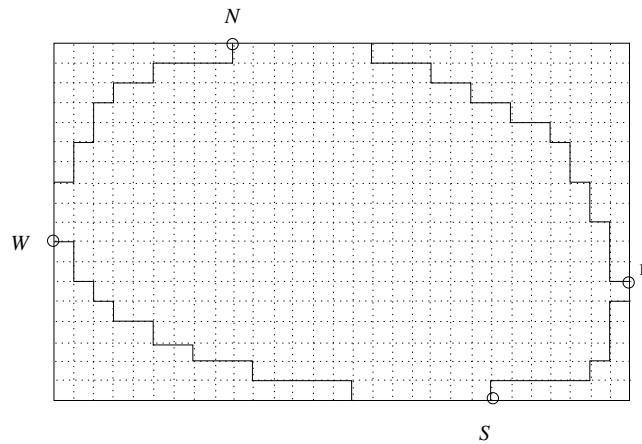


Figure 7.7: A convex polyomino and its standard factorization. The word $w \in \{0, 1\}^*$ coding the WN path is $w = 1110110110100100001$.

Definition 7.3 A word in $\{0, 1\}$ is *WN-convex* if it is the WN path of a convex set.

The *NE*, *ES*, and *SW convexity* can be defined similarly. Obviously, a path is convex if its standard factorization is made by four paths that are *WN*, *NE*, *ES*, and *SW* convex.

7.2.1 Perturbations on the WN paths

From now on, we will consider the WN path only, assuming that all its properties hold for the other three paths up to rotations. In [9], the authors characterized the words that are border of a polyomino by means of Lyndon and Christoffel words:

Property 7.4 A word w is *WN-convex* if and only if its unique Lyndon factorization $w_1^{n_1} w_2^{n_2} \dots w_k^{n_k}$ is such that all w_i are primitive Christoffel words.

Example 83 Consider the following WN -convex path $v = 1011010100010$. The Lyndon factorization of v represented in Figure 7.8 is given by:

$$v = (1)^1(011)^1(01)^2(0001)^1(0)^1;$$

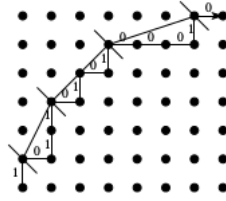


Figure 7.8: Example of a Digitally convex path formed of Christoffel words using the Lyndon factorization.

where 1, 011, 01, 0001 and 0 are all Christoffel words. The Christoffel words are arranged in a weakly decreasing order of slopes:

$$\frac{1}{0} > \frac{2}{1} > \frac{1}{1} > \frac{1}{3} > \frac{0}{1}.$$

Such a result highlights the fact that a WN convex path is composed by line segments, i.e. Christoffel words, having a decreasing slope, so that they respect the lexicographical order and produce a Lyndon factorization. Furthermore, in the same paper, the authors pointed out that such a decomposition can be obtained in linear time.

In this chapter, we will see the importance and impact of the furthest point Q on the Christoffel path with respect to the line segment as mentioned in Chapters 1 and 2. , Therefore, let us consider a primitive Christoffel word w and define $min(w)$ to be the length of the prefix needed to reach the minimal point Q .

As an example, let us consider the primitive Christoffel word $w = 00100101$ be a primitive Christoffel word; its minimal point is at position $(4, 1)$ reached by the prefix 00100 of w , so $min(w) = |00100| = 5$. Since we assume w to be primitive, then $min(w)$ is unique. Figure 7.9 shows a WN path and the minimal points of the Christoffel words.

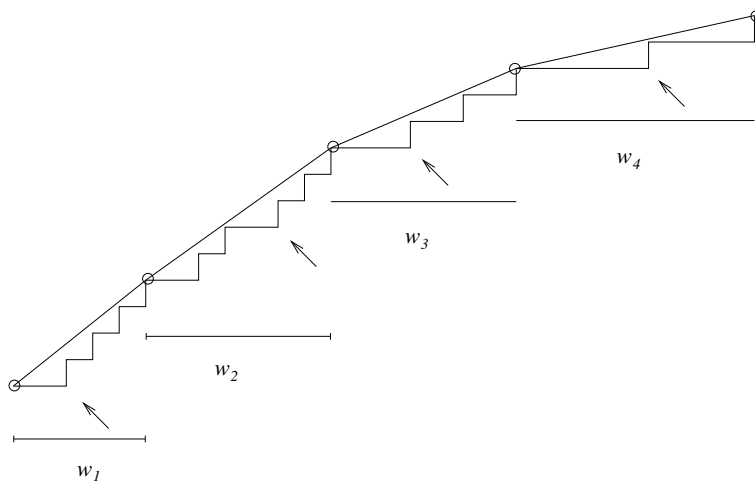


Figure 7.9: A WN path and its decomposition into four Christoffel words w_1 , w_2 , w_3 , and w_4 related to four line segments. The four minimal points of each segment are highlighted.

We remark that, by definition, if $k = \min(w)$, then $w[k] = 0$, and $w[k+1] = 1$. The following property states that if we flip the two letters of w at positions k and $k+1$, we obtain a new word w' that is not a Christoffel word. On the other hand, this new word w' is the concatenation of two Christoffel words $w'[1 \cdots k]$ and $w'[k+1 \cdots n]$, with $n = |w'| = |w|$. Furthermore, k is the only position allowing such type of decomposition:

Proposition 7.5 *Let w be a primitive Christoffel word of length n and $k = \min(w)$.*

- (i) *The words $u = w[1 \cdots k-1]1$ and $v = 0w[k+2 \cdots n]$, are two Christoffel words;*
- (ii) *for each k' different from k , the words $u' = w[1 \cdots k'-1]1$ and $v' = 0w[k'+2 \cdots n]$ are not both Christoffel words.*

Proof. Existence In Chapter 1 and from Property 1.10, we can write $w = 0w'1$ where w' is a palindrome, and using the standard factorization, $w = 0w_110w_21$ where w_1, w_2 are palindromes and $w[1 \cdots m] = 0w_11$. Without loss of generality, we consider $|w_1| < |w_2|$, then using the symmetric standard factorization defined in Lemma 5.4, $w = 0(w_1, q, w_1)_{ss}1$ where q is a palindrome and can be written as; $q = 1q'1$ since $w[1 \cdots m] = 0w_11$ and q' is exactly the path joining the closest point to the line segment at position m to the furthest point at the line segment at position k .

In addition, we have that $w[m+1] = 0$, therefore we deduce that $q = 10q''01$ and $w = 0(w_1, 10q''01, w_1)_{ss}1$. With this factorization, we get $w_2 = q''01w_1$ which is a palindrome, hence we have $q''01w_1 = w_110q''$. By taking $u_1 = w[1 \cdots k-1]1 = 0w_110q''1$ and $u_2 = 0w[k+1 \cdots n] = 0w_11$, we get the two Christoffel words since by construction we have w_110q'' and w_1 palindromes.

Uniqueness Suppose k' is another position where we can exchange 01 into 10 and we will show that in this case $u'_1 = w[1, k'-1]1$ and $u'_2 = 0w[k'+1 \cdots n]$ are both not Christoffel Words.

We have 3 cases to study, either $k' < m$, or, $m < k' < n$, or $k' > n$

1. If $k' < m$:

If k' is a position in the factor $0w_11$, we have $w_1 = f_101f_2 = f_210f_1$ since it is a palindrome. By letting k' be the position between the 0 and 1, we get: $u'_1 = w[1 \cdots k'-1]1 = 0f_11$ and $u'_2 = 0w[k'+1 \cdots n] = 0f_210q''01w_11$. We have that $0w_21 = 0q''01w_11$ where $q''01w_1$ is a palindrome. Therefore, if u'_2 is a Christoffel Word, we must have: $f_210q''01w_1$ a palindrome, this leads to a contradiction even if f_1 and f_2 are palindromes.

2. If $m < k' < k$:

In this case the position k' is in the factor q'' who is a palindrome and we write: $q'' = s_1 0 1 s_2 = s_2 1 0 s_1$, where k' is exactly the position between 0 and 1. Hence $u'_1 = w[1 \cdots k' - 1]1 = 0w_1 1 0 s_1$ and $u'_2 = 0w[k' + 1 \cdots l] = 0s_2 0 1 w_1 1$. We have $w_2 = s_1 0 1 s_2 0 1 w_1$ which is a palindrome hence $s_2 0 1 w_1$ can't be a palindrome then u_2 is not a Christoffel Word.

3. If $k' > k$:

Same reasoning as the first case.

□

Sage code We give in this part the Sage code used to determine the position of the point Q along the Christoffel path of slope a/b :

```

1 def split(a,b):
2     n=a+b
3     k=0
4     for i in range(n):
5         if a*i % n == n-1:
6             k=i
7     return(k)

```

Listing 7.1: Code of the split operator.

Example 84 Let $w = C(\frac{5}{8})$, using the code 8, we can determine the position of Q as we can in Figure 7.10.

```

1 sage: split(5,8)
2 5

```

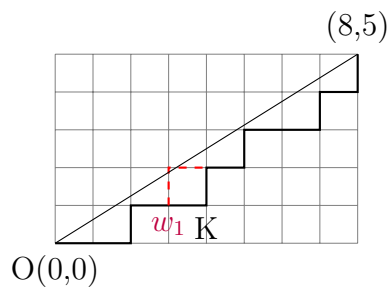


Figure 7.10: Example of using the split operator on the Christoffel word of slope $5/8$.

We point out the following immediate and useful consequence:

Corollary 7.6 *Let w , u and v be as in Property 7.5. It holds*

(i) $\rho(u) > \rho(v)$;

(ii) $\rho(w[1 \cdots k'] 1) > \rho(u)$, for each $k' \neq k (= \min(w))$, and $w[k'] = 0$. A symmetric result holds for v .

7.2.2 Definition of the *split* operator

Proposition 7.5 allows us to define a *split operator*, that acts on the Christoffel word w and provides as output the concatenation of the two words u and v , by simply flipping the sequence 01 at position k and $k + 1$ of w into the sequence 10, i.e., $split(w) = uv$. From now on, we consider the extension of the operator to *sequences* of Christoffel words, and we index it with the (index of the) factor where the split takes place, i.e., if $w = w_1 w_2 \dots w_n$ is a sequence of primitive Christoffel words, then $split_k(w) = w_1 w_2 \dots split(w_k) \dots w_n$. Consecutive applications of the split operator to the word w will be indexed by the sequence of the indexes of the involved factors.

Given two words p_1 and p_2 of the same length l , we say that p_1 is *greater* than or equal to p_2 ($p_1 \geq p_2$), if for each $k \leq l$ it holds $|p_1[1 \cdots k]|_1 \geq |p_2[1 \cdots k]|_1$. The “ \geq ” relation is a natural partial ordering on words.

As an immediate consequence we have:

Property 7.7 *Let $w = w_1 w_2 \dots w_n$ be a sequence of Christoffel words. It holds:*

(i) $split_k(w) \geq w$, with $k \leq n$;

(ii) the split operator commutes with respect to successive applications, i.e.,

$$split_{k,h}(w) = split_{h,k}(w).$$

Attention must be paid when we are dealing with sequences of Christoffel words that are paths of a convex polyomino: the split operator provides an efficient way to add one point on a line segment of the border of the polyomino without loosing the convexity on that segment, but it does not guarantee either to preserve the Lyndon factorization of the related word or its convexity. We can classify the perturbations performed by the split operator on the factor w_i (i.e. $split(w_i) = u_i v_i$) of the Lyndon decomposition of a convex path into three different types, according to the values of the slopes of the consecutive Lyndon factors after the perturbation:

- (a) the Lyndon factorization and the global convexity are preserved (see Fig. 7.11 (a)), i.e., the two new factors u_i and v_i globally preserve the slope decreasing of the line segments of the path;
- (b) the Lyndon factorization is not preserved but the obtained path is still convex (see Fig. 7.11 (b)), i.e., $w_{i-1}u_iv_iw_{i+1}$, with w_{i+1} eventually void, is not a Lyndon factorization, so it does not preserve the slope decreasing of the line segments of the path, while the new Lyndon factorization does;
- (c) neither the Lyndon factorization nor the convexity are preserved (see Fig. 7.11 (c)), i.e., $w_{i-1}u_iv_iw_{i+1}$, with w_{i+1} eventually void, is not a Lyndon factorization. Furthermore, the new Lyndon factorization is not composed by Christoffel words only.

7.2.3 Commutativity of the *split* operator

In what follow, we are interested in showing under which assumptions the commutative behavior of the split operator is preserved in a WN path (as already underlined, by symmetry the found results hold for the remaining three kinds of paths). In particular, we are going to show that, if the split operator produces perturbations of type (a) on two consecutive Christoffel words in a same octant of a WN path, then the result of the two successive applications is independent from their appliance order.

Theorem 7.8 *Let w_1 and w_2 be two consecutive Christoffel words in the same octant of a WN path of a polyomino, and let $\text{split}(w_1) = u_1 v_1$ and $\text{split}(w_2) = u_2 v_2$. If $\rho(v_1) > \rho(w_2)$ and $\rho(w_1) > \rho(u_2)$ (i.e. the split operator provides two perturbations of type (a) on w_1 and w_2), then it holds $\rho(v_1) > \rho(u_2)$.*

Proof. Let $\rho(u_2) = \frac{b}{a}$ and $\rho(v_1) = \frac{b'}{a'}$, and assume without loss of generality that w_1 and w_2 lie in the upper octant, i.e., $a > b$, and $a' > b'$. Since the split operator acts on $\min(w_1)$ and $\min(w_2)$, then it holds

$$\frac{b-1}{a-1} < \rho(w_2) < \frac{b}{a}, \text{ and } \frac{b'-1}{a'-1} < \rho(w_1) < \frac{b'}{a'}. \quad (7.1)$$

Let us proceed by contradiction, assuming that $\rho(v_1) < \rho(u_2)$, i.e., $ab' < ba'$. We first prove that the inequality

$$\frac{b'-1}{a'-1} < \frac{b-1}{a-1} \quad (7.2)$$

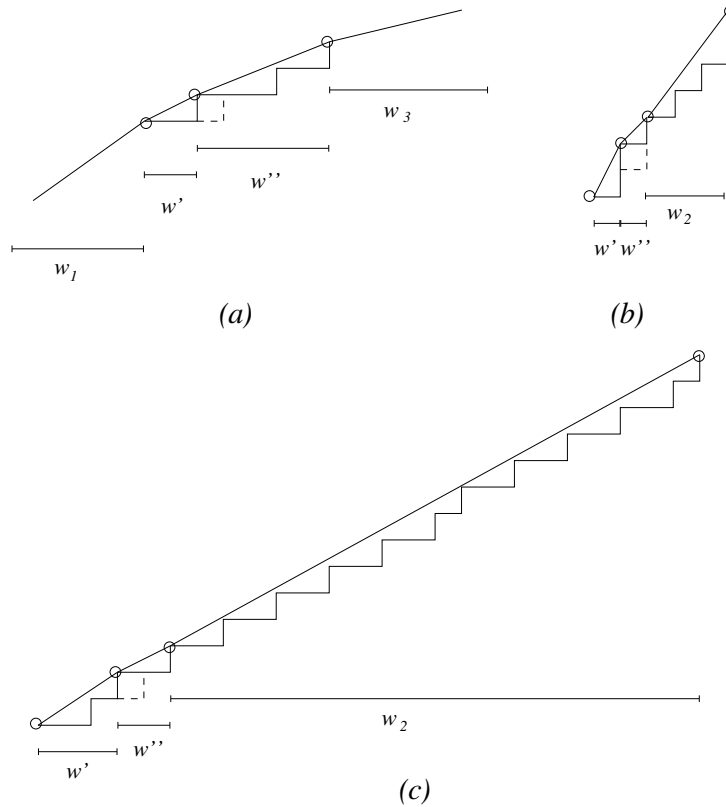


Figure 7.11: Three WN paths of a convex polyomino: (a) the Lyndon factorization and the global convexity are preserved; (b), the global convexity is preserved, but not the Lyndon factorization that has to be modified including w'' in w_2 , i.e., $w'_2 = w'' w_2 = (01)(0101011)$; (c) the global WN convexity is lost. Observe that the word $w'' w_2 = (001)(0010010010010010100100100100101)$ is not a Christoffel word. So, we need a second split in w_2 and the addition of a further point to obtain a WN path back.

is always satisfied: several cases have to be considered according to the mutual dimensions of the four parameters a , a' , b , and b' :

Case 1) $a = a'$. From $a = a'$ it follows $b' < b$ and consequently $b' - 1 < b - 1$, so Inequality 7.2 holds. The case $b = b'$ is symmetrical.

Case 2) $a < a'$. Two sub-cases arise: if $b' < b$, then we have $b' - 1 < b - 1$. Since we also have $a - 1 < a' - 1$, then Inequality 7.2 again is a direct consequence. On the other hand, let $b' > b$. In this case we show that a contradiction arises, i.e., $\rho(w_1) < \rho(u_2)$, against the hypothesis.

The Christoffel tree is isomorphic to the Stern-Brocot tree that contains all the irreducible fractions. The fractions are distributed all over the tree using the Farey addition, which is:

$$\frac{a}{b} \oplus \frac{c}{d} = \frac{a+c}{b+d}.$$

Let the two Christoffel words w_1 and w_2 , of slopes respectively $\rho(w_1)$ and $\rho(w_2)$, be split into Christoffel factors as $w_1 = u_1 v_1$ and $w_2 = u_2 v_2$, as shown in Fig. 7.12.

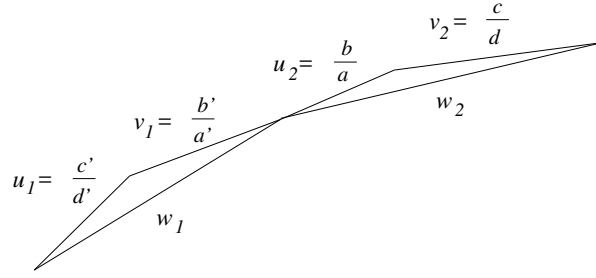


Figure 7.12: The split of the Christoffel words w_1 and w_2 into (u_1, v_1) and (u_2, v_2) .

We let, $\rho(u_1) = \frac{c'}{d'}$, $\rho(v_1) = \frac{b'}{a'}$, $\rho(u_2) = \frac{b}{a}$, and $\rho(v_2) = \frac{c}{d}$.

Using the construction of the Stern-Brocot tree, we know that there exist $k, t \in \mathbb{N}$ such that $c' = b' - k$ and $d' = a' - t$.

Since, by assumption, $b' > b$ and $a' > a$, there also exist $k', t' \in \mathbb{N}$ such that $b = b' - k'$ and $a = a' - t'$.

The following inequalities hold:

$$\begin{aligned} \frac{b'}{a'} &< \frac{b}{a} \\ ab' &< a'b \\ a'b' - a'b &< a'b' - ab' \\ \frac{b' - b}{a' - a} &< \frac{b'}{a'} \\ \frac{k'}{t'} &< \frac{b'}{a'} \\ 2a'k' - 2b't' &< 0. \end{aligned}$$

Reminding that we assumed to be confined in the first octant where the slopes of the Christoffel words are less than 1, then $b < a$ implies $b' - k' < a' - t'$.

Therefore: $-(a' - t') < -(b' - k')$ and, since by the Stern-Brocot tree we have $\frac{c'}{a'} \oplus \frac{k}{t} = \frac{b'}{a'}$, then $\frac{k}{t} < 1$ and consequently $k < t$, then we get

$$-k(a' - t') < -t(b' - k') \text{ i.e. } -k(a' - t') + t(b' - k') < 0.$$

The inequalities gathered up to now lead to the following ones: $\rho(w_1) = \frac{b'+c'}{a'+d'} = \frac{2b'-k}{2a'-t}$ and $\rho(u_2) = \frac{b}{a} = \frac{b'-k'}{a'-t'}$ that are enough to prove that $\rho(w_1) < \rho(u_2)$ always holds against the hypothesis. In fact

$$\begin{aligned} 2b'a' - 2b't' - ka' + kt' &< 2a'b' - 2a'k' - b't + k't \\ 2a'k' - 2b't' - k(a' - t') + t(b' - k') &< 0 \end{aligned}$$

which is always true since $2a'k' - 2b't' < 0$ and $-k(a' - t') + t(b' - k') < 0$.

Case 3) $a > a'$. As above, two sub-cases arise:

- if $b < b'$, then since $\frac{1}{a} < \frac{1}{a'}$ we get $\frac{b}{a} < \frac{b'}{a'}$, i.e., a contradiction to the initial hypothesis.
- The other case concerns $b > b'$. Let us consider $a = a' + h$ and $b = b' + k$ (see Fig. 7.13), and consequently Inequality 7.2 can be written as

$$\frac{b' + k - 1}{a' + h - 1} - \frac{b' - 1}{a' - 1} > 0. \quad (7.3)$$

Since by hypothesis, we have $\frac{b'}{a'} = \frac{b-k}{a-h} < \frac{b}{a}$, hence $ab - ak - ab + hb < 0$ and $\frac{b}{a} < \frac{k}{h}$ as shown in Fig. 7.13. Therefore the following relations hold:

$$\frac{k}{h} > \frac{b}{a} > \frac{b'}{a'} > \frac{b' - 1}{a' - 1}.$$

From the last inequality, it holds that $k(a' - 1) > h(b' - 1)$ and consequently $ka' - hb' - k + h > 0$ that is equivalent to Inequality 7.3.

In the 3 cases, and assuming that $\rho(v_1) > \rho(u_2)$, we obtain Inequality 7.2. The following inequalities are deduced:

$$\frac{b' - 1}{a' - 1} < \frac{b - 1}{a - 1} < \rho(w_2) < \frac{b'}{a'} < \frac{b}{a} < \rho(w_1)$$

and, comparing with the first inequality of the second chain in Inequality 7.1, we get a contradiction. \square

Corollary 7.9 *After performing a sequence of perturbations of type (a) in a WN path, then the obtained path is still a WN path.*

This last result can be rephrased by saying that the split operator commutes in case of perturbations of type (a) inside the same octant.

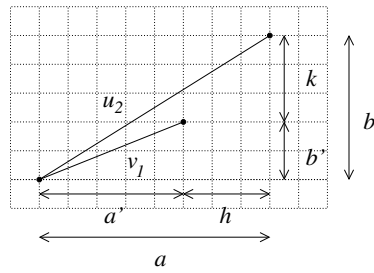


Figure 7.13: An example of the case $a > a'$ and $b > b'$ in the proof of Theorem 7.8.

Open questions A further analysis has to be carried on in presence of perturbations of type (b), both in the same octant and in the whole quadrant. Hence some open questions can be asked at this step of work and we mention some of them:

1. What will happen with the case in which $\text{slope}(v_1) < \text{slope}(w_2)$.
2. The results is still the same if we extend the work to the case of words from different octant.
3. Are we able to produce a final reconstruction algorithm that extends the strategy of Del Lungo-Nivat et al. for HV convex polyominoes by preserving step by step digital convexity.

List of Figures

1.1	Beginning of a standard Sturmian word with an irrational slope. . .	13
1.2	The Lower and Upper Christoffel paths of slope $5/8$ respectively. . .	16
1.3	Illustration of the geometrical definition of Christoffel words. The Christoffel path goes from $(0,0)$ to $(5,3)$ and $C(\frac{3}{5}) = 00100101$. . .	16
1.4	Illustration of the geometrical definition of Upper Christoffel words. The Christoffel path goes from $(0,0)$ to $(3,5)$ and $C(\frac{5}{3}) = 11011010$. . .	17
1.5	The standard factorization at the point C	19
1.6	The two triangles OPA and OCA	20
1.7	The standard factorization of the Christoffel word of slope $5/8$	20
1.8	The words, w, u and v are Christoffel words with palindromic central parts w_1, u_1, v_1 resp. The property of palindromes gives the Palindromic factorization of w denoted by $w = (p_1, p_2)$	21
1.9	The rule followed to construct the Christoffel tree.	22
1.10	The first levels of the Christoffel tree.	23
1.11	The Christoffel word of slope $\frac{6}{10}$	34
2.1	The Christoffel word of slope $5/8$	38
2.2	The labels assigned to each integer point of the Christoffel path of slope $5/8$	39
2.3	The Cayley graph of $w = 00100101$	41
2.4	The representation of w on a line segment with the position angle of each letter.	43
2.5	The axis of symmetry of $w = C(\frac{7}{5})$ represented on the Cayley graph before and after rotation.	43
2.6	Position of $w = C(\frac{3}{7}) = 0001001001$ on the Christoffel tree using $\Delta(\frac{3}{7}) = 0011$	45
2.7	The first levels of the Stern-Brocot tree.	48
2.8	Relation between the Farey sum and the endomorphisms ϕ_0 and ϕ_1	51
2.9	Position of the $w = C(\frac{5}{7})$ comparing to the position of $5/7$ on Stern-Brocot tree.	54

2.10	Illustration of Definition 2.28 in the Stern-Brocot tree relatively to $3/5 = [0, 1, 1, 2]$. We have : TBF(3/5) = 2/3, FRF(3/5) = 1/2, FEF(3/5) = 4/7, FDF(3/5) = 5/8, FPF(3/5) = 1/3 and SUF(3/5) = 1/1.	59
2.11	Form of the Christoffel word $C(\frac{a}{b})$ obtained by ϕ_1	62
2.12	Form of the Christoffel word $C(\frac{a}{b})$ obtained by ϕ_0	62
3.1	The first three levels of balancedness for the Christoffel word of slope 3/5 representation on a cayley graph.	70
3.2	The general form of the matrix $U_{\frac{1}{b}}$	77
3.3	Position of the fractions $\frac{u}{v}$, $\frac{x}{y}$, $\frac{c}{d}$, $\frac{e}{f}$, $\frac{g}{h}$, $\frac{p}{q}$ and $\frac{s}{t}$ relatively to $\frac{a}{b}$, for the case where $a_{z-1} \geq 2$ and z odd. By definition, we have : $\frac{a}{b} = [a_0 \dots, a_z]$, $\frac{u}{v} = [a_0 \dots, a_{z-1} + 1]$, $\frac{x}{y} = [a_0, \dots, a_{z-1}]$, $\frac{p}{q} = [a_0, \dots, a_{z-2} + 1]$, $\frac{s}{t} = [a_0, \dots, a_{z-2}]$, $\frac{e}{f} = [a_0, \dots, a_{z-1} - 1, 2]$, $\frac{c}{d} = [a_0, \dots, a_{z-1}, 2]$, $\frac{g}{h} = [a_0, \dots, a_{z-1} - 1]$	78
3.4	The decomposition of the matrix $U_{\frac{5}{8}}$ into 9 blocks. We have TBF($\frac{5}{8}$) = $\frac{3}{5}$, SUF($\frac{5}{8}$) = $\frac{1}{2}$, FPF($\frac{5}{8}$) = $\frac{3}{4}$, $\frac{x}{y} = \frac{2}{3}$ and $k = 5$	80
4.1	The first four levels of the Stern-Brocot tree containing the fractions and their second order balance value.	90
4.2	Case where $a_z \geq 4$ and $a_z = 3$ respectively.	94
4.3	Geometric representation of the path having $a_z = 2$ and $a_{z-1} = 1$	94
4.4	Geometric representation of the path having $a_z = 2$ and $a_{z-1} = 2$	95
4.5	Geometric representation for the case $a_{z-1} = 4$	96
4.6	Geometric representation for the path $a_{z-1} = 3$	96
4.7	The Fibonacci sequence in the Stern-Brocot tree.	98
5.1	The symmetric standard factorization of $w = C(\frac{5}{8})$	113
5.2	The geometrical representation of $\Gamma(\frac{1}{6})$ joining the closest and furthest points of the Christoffel path.	114
5.3	The symmetric standard factorization of the Christoffel words of slopes $x/y, u/v$ and c/d with an odd value for a_{z-1}	116
5.4	The symmetric standard factorization of the Christoffel words of slopes $x/y, u/v$ and c/d with an odd value for a_{z-2} and $a_{z-1} = 2$	117
5.5	The symmetric standard factorization of the Christoffel words of slopes $x/y, u/v$ and c/d with an even value for a_{z-1}	117
5.6	The symmetric standard factorization of the Christoffel words of slopes $x/y, u/v$ and c/d with an even value for a_{z-2} and $a_{z-1} = 2$	118
5.7	The geometric representation of the α, γ and β -blocks of $U_{\frac{3}{11}}$	122
5.8	The geometric representation of the α, γ and β -blocks of $U_{\frac{6}{11}}$	125

6.1	Two representation of the Christoffel word 11211212 of slope $3/5$. On the left as a lattice path and on the right as a Cayley graphs. The representation as Cayley graphs highlights the fact that the Christoffel word of slope p/q has the same structure then the one of slope q/p and that both synchronizes into a circular conjugate. .	129
6.2	The vertical invariant of the synchronized form of the generators $g_1 = 3$ and $g_2 = 5$	133
6.3	The three matrices $U_{\frac{1}{6}}$, $U_{\frac{2}{5}}$ and $U_{\frac{3}{4}}$ used for the synchronization of the triplet $(1, 2, 4) \bmod 7$	154
6.4	The two matrices $U_{\frac{3}{8}}$ and $U_{\frac{5}{6}}$ used for the synchronization of the triplet $(3, 3, 5) \bmod 11$	155
7.1	The left square shows the 4 possible directions of a point in a 4- connected path and the right square shows the 8 possible directions of a point in an 8-connected path.	160
7.2	Example of an MP-convex and H-convex.	161
7.3	Example of an MP-convex but not H-convex.	161
7.4	An example of a polyomino with border word: $w = Bd(P) =$ $010\overline{1001110011001}$	162
7.5	A path coded by the word: $w = 1111\overline{00111000}$	162
7.6	An example of a polyomino in the \mathbb{Z}^2 lattice.	163
7.7	A convex polyomino and its standard factorization. The word $w \in$ $\{0, 1\}^*$ coding the WN path is $w = 1110110110100100001$	164
7.8	Example of a Digitally convex path formed of Christoffel words using the Lyndon factorization.	165
7.9	A WN path and its decomposition into four Christoffel words w_1 , w_2 , w_3 , and w_4 related to four line segments. The four minimal points of each segment are highlighted.	165
7.10	Example of using the split operator on the Christoffel word of slope $5/8$	167
7.11	Three WN paths of a convex polyomino: (a) the Lyndon factoriza- tion and the global convexity are preserved; (b), the global convexity is preserved, but not the Lyndon factorization that has to be mod- ified including w'' in w_2 , i.e., $w'_2 = w'' w_2 = (01)(0101011)$; (c) the global WN convexity is lost. Observe that the word $w''w_2 =$ $(001)(0010010010010010100100100101)$ is not a Christoffel word. So, we need a second split in w_2 and the addition of a further point to obtain a WN path back.	170
7.12	The split of the Christoffel words w_1 and w_2 into (u_1, v_1) and (u_2, v_2) .171	
7.13	An example of the case $a > a'$ and $b > b'$ in the proof of Theorem 7.8.173	

List of Tables

1.1	The right and left factorization of $w_1 = xyxxzxy$, $w_2 = xxyy$, and $w_3 = xyxxyxy$	12
1.2	The Lyndon factorization of $w_1 = 100101100101010$, $w_2 = 0010010$ and $w_3 = 001010110010$	13
1.3	The lexicographic order of all the conjugates of $w = C(\frac{2}{3}) = 00101$	18
1.4	The circular factors of the word: $w = 00101$	23
4.1	The representation of the 4 cases for $a_{z-1} = 1$	94
4.2	The representation of the 2 cases for $a_{z-1} = 2$	95
4.3	The representation of the 2 cases for $a_{z-1} = 3$	96
4.4	Decomposition of an element of V_i^j by factors of length j	102
4.5	Decomposition of the element $v_i^{j-\gamma+1}$ by factors of length j	102
4.6	Table of Parikh vectors of length 5 for all the factors of $w = C(\frac{4}{9})$ of length 9.	108

Listings

1.1	Code of the Balance Matrix.	28
2.1	Coding a word on the Cayley graph with its center of symmetry. . .	44
2.2	Coding used to get the elements of the Stern-Brocot tree represented by couples.	52
3.1	Code of the Second order balance Matrix.	72
4.1	Code of the paths that minimize the growth of δ^2 on the Stern- Brocot tree.	99
4.2	Minimal value of δ^2 for a fixed length.	100
4.3	Code of the matrix AB	106
7.1	Code of the split operator.	167

Index

iterated palindromic closure, 53
alphabet, 9, 11–13, 15, 16, 23, 27, 37
abelian vector, 105
balance matrix, 23–25, 27, 29, 31–33, 69
coprime, 14, 15, 19, 34
cayley graph, 41, 128
circular word, 10, 23, 24, 30
complementary, 29, 32, 73
conjugate, 10, 11, 18, 21
christoffel tree, 21, 22, 72
Christoffel words, 11, 15, 17, 18, 20, 22, 37, 163
digital convexity, 157
discrete set, 159
directive sequence, 44, 52
factor, 10, 24–26
Freeman Chain code, 161
first deviation fraction, 58
first extended fraction, 58, 95
Fibonacci sequence, 97
first parallel fraction, 58
first reduced fraction, 58
Farey Sum, 48, 72, 170
generator, 128
hole, 158
language, 23, 24
length, 9, 14, 17, 24–26, 29–31
lexicographic order, 10–12, 18, 25, 165
lyndon factorization, 12, 13, 163, 165
lyndon words, 11–13, 18, 25
palindrome, 10, 17, 20, 21
pick's theorem, 19

period, 14, 26, 30
 palindrome factorization, 39
 polyominoes, 158, 163
 partial denominator, 47, 93
 prefix, 10, 12, 25
 proper, 11, 12, 17, 18
 primitive, 9, 11, 17, 24
 Parikh vector, 105, 106
 regular, 12
 reversal, 10
 Stern-Brocot tree, 48, 72, 98, 171
 seed, 128
 standard factorization, 12, 18–22, 39, 51, 166
 synchronized rows, 75, 79
 second order balance matrix, 71, 105
 symmetric standard factorization, 111, 166
 standard sturmian, 13, 14
 second unidirectional father, 58
 suffix, 10, 12, 25
 synchronization, 128
 top branch fraction, 58
 zigzag path, 92, 98
Pal, 53
 \perp , 15, 34
 Δ , 44
FDF, 58
FEF, 58
FPF, 58
FRF, 58
 $|w|$, 9, 14, 17, 24, 26, 30, 31
 $\langle \iota$, 11
SUF, 58
TBF, 58

Bibliography

- [1] B. Achille. Calcul des rouages par approximation. *Revue chronométrique. Journal des horlogers, scientifique et pratique*, 3, 1861.
- [2] E. Altman, B. Gaujal, and A Hordijk. Balanced sequences and optimal routing. *Journal of the ACM*, 47(4):752–775, 2000.
- [3] J. Berstel. Tracé de droites, fractions continues et morphismes itérés, 1990.
- [4] J. Berstel and A. De Luca. Sturmian words, lyndon words and trees. *Theoretical Computer Science*, 178(1):171–203, 1997.
- [5] J. Berstel, A. Lauve, C. Reutenauer, and F. Saliola. Combinatorics on words: Christoffel words and repetition in words. 2008.
- [6] V. Berthé, A. De Luca, and C. Reutenauer. On an involution of christoffel words and sturmian morphisms. *European Journal of Combinatorics*, 29(2):535–553, 2008.
- [7] J-P. Borel and F. Laubie. Quelques mots sur la droite projective réelle. *Journal de théorie des nombres de Bordeaux*, 5(1):23–51, 1993.
- [8] J-P. Borel and C. Reutenauer. On christoffel classes. *RAIRO-Theoretical Informatics and applications*, 40(1):15–27, 2006.
- [9] S. Brlek, J-O. Lachaud, X. Provençal, and C. Reutenauer. Lyndon + christoffel=digitally convex. *Pattern Recognition*, pages 42–10, 2009.
- [10] G. Castiglione, A. Frosini, E. Munarini, A. Restivo, and S. Rinaldi. Combinatorial aspects of l-convex polyominoes. *European Journal of Combinatorics*, 28(6):1724–1741, 2007.
- [11] G. Castiglione, A. Frosini, A. Restivo, and S. Rinaldi. Enumeration of l-convex polyominoes by rows and columns. *Theoretical Computer Science*, 347(1-2):336–352, 2005.

- [12] B. Chaudhuri and A. Rosenfeld. On the computation of the digital convex hull and circular hull of a digital region. *Pattern Recognition*, 31 (12), 1998.
- [13] K. Chen, R. Fox, and R. Lyndon. Free differential calculus, iv. the quotient groups of the lower central series. *Annals of Mathematics*, pages 81–95, 1958.
- [14] E. Christoffel. Observatio arithmetica. *Annali di Matematica Pura ed Applicata (1867-1897)*, 6(1):148–152, 1875.
- [15] W-F. Chuan. α -words and factors of characteristic sequences. *Discrete Mathematics*, 177(1-3):33–50, 1997.
- [16] F. De Carli, A. Frosini, S. Rinaldi, and L. Vuillon. How to construct convex polyominoes on dna wang tiles. *LAMA report, Lama-Univ. Savoie, France*, 2009.
- [17] A. de Luca and F. Mignosi. Some combinatorial properties of sturmian words. *Theoretical Computer Science*, 136(2):361–385, 1994.
- [18] A. Del Lungo, E. Duchi, A. Frosini, and S. Rinaldi. Enumeration of convex polyominoes using the eco method. In *DMCS*, pages 103–116, 2003.
- [19] A. Del Lungo, E. Duchi, A. Frosini, and S. Rinaldi. On the generation and enumeration of some classes of convex polyominoes. *the electronic journal of combinatorics*, 11(1):R60, 2004.
- [20] P. Dulio, A. Frosini, S. Rinaldi, L. Tarsissi, and L. Vuillon. First steps in the algorithmic reconstruction of digital convex sets. In *International Conference on Combinatorics on Words*, pages 164–176. Springer, 2017.
- [21] S. Dulucq and D. Gouyou-Beauchamps. Sur les facteurs des suites de sturm. *Theoretical Computer Science*, 71(3):381–400, 1990.
- [22] J. Duval. Mots de lyndon et périodicité. *RAIRO, Informatique théorique*, 14(2):181–191, 1980.
- [23] N-J. Fine and H-S. Wilf. Uniqueness theorems for periodic functions. *Proceedings of the American Mathematical Society*, 16(1):109–114, 1965.
- [24] A. Fraenkel. The bracket function and complementary sets of integers. *Canadian Journal of Mathematics. Journal Canadien de Mathématiques*, 21:6–27, 1969.
- [25] A. Fraenkel, J. Levitt, and M. Shimshoni. Characterization of the set of values $f(n)=[na]$, $n= 1, 2, \dots$. *Discrete Mathematics*, 2(4):335–345, 1972.

- [26] H. Freeman. On the encoding of arbitrary geometric configurations. *IRE Transactions on Electronic Computers*, (2):260–268, 1961.
- [27] M. Gardner. Mathematical games. *Scientific American*, 103, 1958.
- [28] S-W. Golomb. *Polyominoes: puzzles, patterns, problems, and packings*. Princeton University Press, 1996.
- [29] R-E. Graham, D-E. Knuth, and O. Patashnik. Concrete mathematics. a foundation for computer science (second ed.). *Addison-Wesley, Reading, MA*, 1994.
- [30] R-L. Graham. Covering the positive integers by disjoint sets of the form $\{[n\alpha + \beta]: n= 1, 2, \dots\}$. *Journal of Combinatorial Theory, series A*, 15(3):354–358, 1973.
- [31] G-H. Hardy and E-M. Wright. *An introduction to the theory of numbers*. Oxford University Press, 1979.
- [32] P. Hubert. Suites équilibrées. *Theoretical Computer Science*, 242(1):91–108, 2000.
- [33] M. Lothaire. *Algebraic combinatorics on words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2002.
- [34] R. Lyndon. Identities in finite algebras. *Proceedings of the American Mathematical Society*, 5(1):8–9, 1954.
- [35] G. Melançon. Lyndon words and singular factors of sturmian words. *Theoretical computer science*, 218(1):41–59, 1999.
- [36] M. Minsky and S. Papert. *perceptrons*. MIT Press, 1988.
- [37] R. Osborne and H. Zieschang. Primitives in the free group on two generators. *Inventiones mathematicae*, 63(1):17–24, 1981.
- [38] G. Paquin and C. Reutenauer. On the superimposition of christoffel words. *Theoretical Computer Science*, 412(4):402–418, 2011.
- [39] G. Paquin and L. Vuillon. A characterization of balanced episturmian sequences. *Journal of Combinatorics*, 14(2):R33, 2007.
- [40] X. Provençal and L. Vuillon. Discrete segments of z^3 constructed by synchronization of words. *Discrete Applied Mathematics*, 2014.

- [41] G-N. Raney. *On continued fractions and finite automata*. *Mathematische Annalen*, 206(4):265–283, 1973.
- [42] C. Reutenauer. *Free lie algebras*. *Handbook of algebra*, 3:887–903, 2003.
- [43] C. Reutenauer. *From christoffel words to markoff numbers*, 2016.
- [44] G. Richomme, K. Saari, and L-Q. Zamboni. *Abelian properties of words*. arXiv preprint arXiv:0904.2925, 2009.
- [45] A. Shirshov. *Subalgebras of free lie algebras*. *Contemporary Mathematicians*, page 3, 1953.
- [46] J. Simpson. *Disjoint beatty sequences*. *Integers: Electronic Journal of Combinatorial Number Theory*, 4(A12):A12, 2004.
- [47] Th. Skolem. *On certain distributions of integers in pairs with given differences*. *Mathematica Scandinavica*, 5, 1957.
- [48] J. Slansky. *Recognition of convex blobs*. *Pattern Recognition*, 2:3–10, 1970.
- [49] H-J. Smith. *Note on continued fractions*. *Messenger Math*, 6(1-14):3–54, 1876.
- [50] R. Tijdeman. *On complementary triples of sturmian bisequences*. *Indagationes Mathematicae*, 7(3):419–424, 1996.
- [51] R. Tijdeman. *Exact covers of balanced sequences and Fraenkel’s conjecture*. In *Algebraic number theory and Diophantine analysis (Graz, 1998)*, pages 467–483. *de Gruyter, Berlin*, 2000.
- [52] R. Tijdeman. *Fraenkel’s conjecture for six sequences*. *Discrete Mathematics*, 222(1):223–234, 2000.
- [53] L. Vuillon. *Balanced words*. *Bulletin of the Belgian Mathematical Society-Simon Stevin*, 10(5):787–805, 2003.