Contributions to combinatorics on words in an abelian context and covering problems in graphs

Elise Vandomme

UNIVERSITÉ DE GRENOBLE

Université de Liège

THÈSE

Dans le cadre d'une cotutelle entre l'Université de Grenoble et l'Université de Liège pour obtenir les grades de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE et

DOCTEUR EN SCIENCES DE L'UNIVERSITÉ DE LIÈGE

Spécialité : **Mathématiques**

Arrêté ministériel : le 6 janvier 2005 – 7 août 2006

Présentée par **Élise VANDOMME**

Thèse dirigée par **Sylvain GRAVIER** et **Michel RIGO**

Préparée au sein de l'Institut Fourier et du Département de mathématique dans les écoles doctorales ED MSTII et ED de Sciences

Contributions to combinatorics on words in an abelian context and covering problems in graphs

Thèse soutenue publiquement le 7 janvier 2015,
devant le jury composé de :

Pr Pierre LECOMTE, ULg	Président
D.R. Valérie BERTHÉ, CNRS - LIAFA	Rapportrice
Pr Éric SOPENA, LABRI	Rapporteur
Pr Philippe CARA, VUB	Examinateur
Dr Éric DUCHÊNE, IUT Lyon - LIRIS	Examinateur
D.R. Sylvain GRAVIER, CNRS - UJF	Directeur de thèse
Pr Michel RIGO, ULg	Directeur de thèse

Abstract

This thesis dissertation is divided into two (distinct but connected) parts that reflect the joint PhD. We study and we solve several questions regarding on the one hand combinatorics on words in an abelian context and on the other hand covering problems in graphs. Each particular problem is the topic of a chapter.

In combinatorics on words, the first problem considered focuses on the 2-regularity of sequences in the sense of Allouche and Shallit. We prove that a sequence satisfying a certain symmetry property is 2-regular. Then we apply this theorem to show that the 2-abelian complexity functions of the Thue–Morse word and the period-doubling word are 2-regular. The computation and arguments leading to these results fit into a quite general scheme that we hope can be used again to prove additional regularity results.

The second question concerns the notion of return words up to abelian equivalence, introduced by Puzynina and Zamboni. We obtain a characterization of Sturmian words with non-zero intercept in terms of the finiteness of the set of abelian return words to all prefixes. We describe this set of abelian returns for the Fibonacci word but also for the Thue–Morse word (which is not Sturmian). We investigate the relationship existing between the abelian complexity and the finiteness of this set.

In graph theory, the first problem considered deals with identifying codes in graphs. These codes were introduced by Karpovsky, Chakrabarty and Levitin to model fault-diagnosis in multiprocessor systems. The ratio between the optimal size of an identifying code and the optimal size of a fractional relaxation of an identifying code is between 1 and $2\ln(|V|)+1$ where V is the vertex set of the graph. We focus on vertex-transitive graphs, since we can compute the exact fractional solution for them. We exhibit infinite families, called generalized quadrangles, of vertex-transitive graphs with integer and fractional identifying codes of order $|V|^\alpha$ with $\alpha \in \{1/4, 1/3, 2/5\}$.

The second problem concerns (r, a, b)-covering codes of the infinite grid already studied by Axenovich and Puzynina. We introduce the notion of constant 2-labellings of weighted graphs and study them in four particular weighted cycles. We present a method to link these labellings with covering codes. Finally, we determine the precise values of the constants a and b of any (r, a, b)-covering code of the infinite grid with $|a - b| > 4$. This is an extension of a theorem of Axenovich.

Key words: combinatorics on words, ℓ-abelian equivalence, regularity, recurrence, abelian return words, Sturmian words, graph theory, identifying codes, vertex-transitive graphs, generalized quadrangles, (r, a, b)-covering codes, infinite grid.

Bref résumé

Cette dissertation se divise en deux parties, distinctes mais connexes, qui sont le reflet de la cotutelle. Nous étudions et résolvons des problèmes concernant d'une part la combinatoire des mots dans un contexte abélien et d'autre part des problèmes de couverture dans des graphes. Chaque question fait l'objet d'un chapitre.

En combinatoire des mots, le premier problème considéré s'intéresse à la régularité des suites au sens défini par Allouche et Shallit. Nous montrons qu'une suite qui satisfait une certaine propriété de symétrie est 2-régulière. Ensuite, nous appliquons ce théorème pour montrer que les fonctions de complexité 2-abélienne du mot de Thue–Morse ainsi que du mot appelé "period-doubling" sont 2-régulières. Les calculs et arguments développés dans ces démonstrations s'inscrivent dans un schéma plus général que nous espérons pouvoir utiliser à nouveau pour prouver d'autres résultats de régularité.

Le deuxième problème poursuit le développement de la notion de mot de retour abélien introduite par Puzynina et Zamboni. Nous obtenons une caractérisation des mots sturmiens avec un intercepte non nul en termes du cardinal (fini ou non) de l'ensemble des mots de retour abélien par rapport à tous les préfixes. Nous décrivons cet ensemble pour Fibonacci ainsi que pour Thue–Morse (bien que cela ne soit pas un mot sturmien). Nous étudions la relation existante entre la complexité abélienne et le cardinal de cet ensemble.

En théorie des graphes, le premier problème considéré traite des codes identifiants dans les graphes. Ces codes ont été introduits par Karpovsky, Chakrabarty et Levitin pour modéliser un problème de détection de défaillance dans des réseaux multiprocesseurs. Le rapport entre la taille optimale d'un code identifiant et la taille optimale du relâchement fractionnaire d'un code identifiant est comprise entre 1 et $2\ln(|V|) + 1$ où V est l'ensemble des sommets du graphe. Nous nous concentrons sur les graphes sommet-transitifs, car nous pouvons y calculer précisément la solution fractionnaire. Nous exhibons des familles infinies, appelées quadrangles généralisés, de graphes sommet-transitifs pour lesquelles les solutions entière et fractionnaire sont de l'ordre $|V|^\alpha$ avec $\alpha \in \{1/4, 1/3, 2/5\}$.

Le second problème concerne les (r, a, b)-codes couvrants de la grille infinie déjà étudiés par Axenovich et Puzynina. Nous introduisons la notion de 2-coloriages constants de graphes pondérés et nous les étudions dans le cas de quatre cycles pondérés particuliers. Nous présentons une méthode permettant de lier ces 2-coloriages aux codes couvrants. Enfin, nous déterminons les valeurs exactes des constantes a et b de tout (r, a, b)-code couvrant de la grille infinie avec $|a - b| > 4$. Il s'agit d'une extension d'un théorème d'Axenovich.

Mots clés : combinatoire des mots, équivalence ℓ-abélienne, régularité, récurrence, mots de retour abélien, mots sturmiens, théorie des graphes, codes identifiants, graphes sommet-transitifs, quadrangles généralisés, (r, a, b)-codes couvrants, grille infinie.

Remerciements

Un doctorat est un long chemin dont le tracé change en cours de route. Comme l'important n'est pas seulement la destination mais aussi le voyage, je tiens à remercier tous les gens qui ont contribué de près ou de loin à la réalisation de cette thèse.

Je remercie vivement mes deux promoteurs Michel RIGO *et Sylvain* GRAVIER *pour m'avoir permis de mener à bien ce projet de doctorat. Dès ma deuxième année de master, ils m'ont initiés à la recherche et m'ont fait découvrir leurs thèmes de recherches. Merci à Michel pour sa disponibilité, son soutien et ses conseils. Merci à Sylvain pour sa confiance et ses encouragements. Merci à tous les deux de m'avoir donné le gout de la vulgarisation à travers Maths à Modeler.*

J'adresse également mes remerciements à Valérie BERTHÉ*, Philippe* CARA*, Éric* DUCHÊNE*, Pierre* LECOMTE *et Éric* SOPENA *qui me font l'honneur d'être membres de mon jury. Je remercie en particulier Valérie* BERTHÉ *et Éric* SOPENA *d'avoir accepté d'être rapporteurs de cette thèse.*

Lors de ce doctorat, j'ai eu le plaisir de collaborer avec d'autres chercheurs : Anne LACROIX*, Aline* PARREAU*, Narad* RAMPERSAD*, Sara* ROTTEY*, Eric* ROWLAND*, Pavel* SALIMOV *et Leo* STORME*. Je les remercie chaleureusement pour ces collaborations enrichissantes.*

J'ai eu le plaisir d'effectuer ce doctorat en tant qu'assistante au Département de Mathématiques de l'Université de Liège ainsi qu'en tant que membre de l'Institut Fourier de Grenoble. Je remercie tous mes collègues pour la bonne ambiance de travail. En particulier, je tiens à remercier mes "soeurs" de thèse liégeoises et grenobloises pour leur soutien lors des exposés, l'équipe de Mathématiques Discrètes de Liège pour son dynamisme, l'équipe Maths à Modeler de Grenoble pour son accueil, les "matheuses bavardes" pour décompresser.

J'exprime toute ma gratitude à mes collègues, amis et membres de la famille qui ont relu certains passages de ce manuscrit. Enfin, je remercie de tout cœur mes parents et Thomas pour leur soutien inconditionnel et leurs encouragements permanents.

Contents

Aperçu de la thèse

Introduction

Cette thèse traite de deux sujets (distincts mais liés) qui sont la combinatoire des mots et la théorie des graphes. Une fois n'est pas coutume, commençons par une mise en situation de la théorie des graphes. Le jeu "*Qui est-ce ?*" était un jeu très à la mode quand j'étais petite. Il s'agit d'un jeu à deux joueurs où chaque joueur dispose d'un plateau sur lequel sont représentés les portraits de 24 personnages. Au début de la partie, chaque joueur choisit secrètement l'un de ces personnages. Le but du jeu est alors de deviner le personnage choisi par l'adversaire, en posant des questions sur son apparence physique. Par exemple, un joueur peut demander à l'autre si le personnage choisi a des lunettes. Les questions sont posées à tour de rôle. Imaginons une version statique de ce jeu où toutes les questions d'un joueur sont posées d'un coup et regardons un exemple concret (où les nombres de personnages et de caractéristiques sont réduits). Le plateau représenté à la Figure 1 contient six personnages qui ont éventuellement une canne, un chapeau, des lunettes ou une moustache. Combien de questions devons-nous poser pour pouvoir identifier le personnage choisi à coup sûr ?

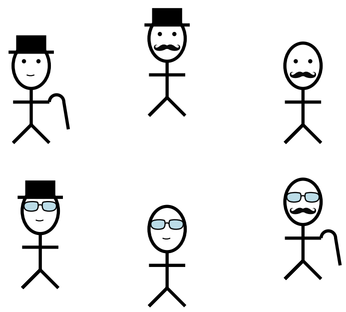


Figure 1: Exemple de plateau de jeu "*Qui est-ce ?*".

Dans cet exemple, il suffit de trois questions à propos de la possession d'une canne, de lunettes et de chapeau, comme le montre la Figure 2. Les trois attributs canne, lunettes et chapeau permettent donc d'identifier le personnage choisi.

Transposons cet exemple au monde de la théorie des graphes. Un graphe est un ensemble de points qui sont reliés entre eux par des lignes, où les points et les lignes sont appelés respectivement sommets et arêtes. Les graphes permettent de modéliser divers problèmes, comme par exemple, un plan schématique du métro dans une ville, un arbre généalogique,

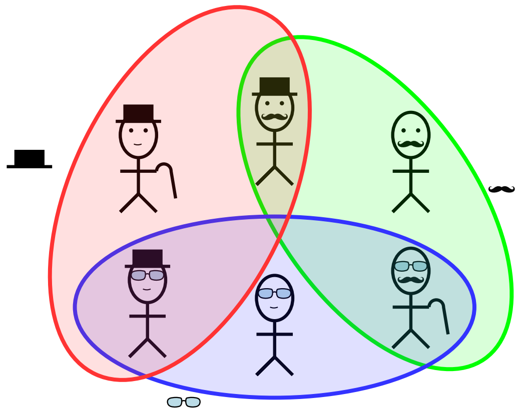


Figure 2: Trois questions suffisent pour identifier le personnage choisi secrètement.

la représentation d'un réseau informatique. Ils peuvent aussi modéliser le plan d'une maison (Figure 3). Les pièces deviennent des sommets et deux sommets sont reliées si les pièces correspondantes sont voisines l'une de l'autre.

Figure 3: Une maison vue de trois façons différentes : en 3 dimensions, en 2 dimensions et modélisée par un graphe. Les lettres S, C et T indiquent l'emplacement du salon, de la cuisine et des toilettes dans chacune des vues.

Les attributs d'un sommet ne sont plus des caractéristiques physiques, mais des caractéristiques liées à la structure du graphe. Pour un sommet fixé, son attribut est l'ensemble des sommets qui lui sont reliés. Reprenons l'exemple de la maison. L'heureux propriétaire de cette maison souhaite la protéger d'un incendie en installant des détecteurs de feu dans certaines pièces. Un détecteur peut détecter un incendie s'il se déclare dans la même pièce que le détecteur ou dans une pièce voisine. La question naturelle est de déterminer un placement de détecteurs qui permettra d'identifier l'endroit exact d'un éventuel incendie.

Plaçons les détecteurs dans les pièces selon le dessin à gauche de la Figure 4. Ces détecteurs permettent de détecter s'il y a un incendie mais ne permettent pas de connaître la position exacte de l'incendie. Par exemple, si le détecteur dans le salon est le seul à s'allumer, le feu peut se trouver dans la cuisine, dans les toilettes ou encore ailleurs. Pour résoudre ce problème, nous pouvons ajouter de nouveaux détecteurs. Plaçons-les comme dans le dessin à droite de la Figure 4. Cet agencement de détecteurs permet bien de signaler si un incendie se déclare. De plus, il n'y a plus d'ambiguïté sur la position de l'incendie. Par exemple, si le détecteur du salon est le seul à signaler l'incendie, alors le feu se trouve dans les toilettes.

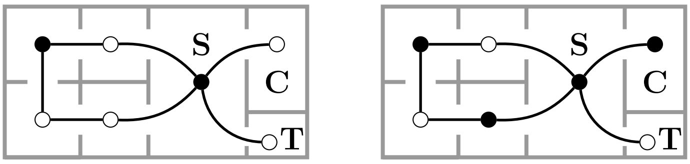


Figure 4: Deux placements de détecteurs d'incendie dans le graphe représentant la maison. Les sommets noirs et les sommets blancs symbolisent respectivement les pièces avec un détecteur et les pièces sans.

L'ensemble des détecteurs (i.e., l'ensemble des sommets noirs) forme un *code identifiant.* Ces codes ont été introduits en 1998 par Karvosky, Chakrabarty et Levitin [KCL98] pour modéliser un problème d'identification de processeurs défectueux dans un réseau multiprocesseur. Plus tard, d'autres applications furent découvertes telles que la conception de réseaux de détecteurs d'incendie dans les bâtiments [UTS04].

La question essentielle dans le cadre des codes identifiants est de déterminer pour un graphe donné, la taille minimale d'un code identifiant de ce graphe (c'est-à-dire le nombre minimum de sommets qui composent le code identifiant). Dans l'exemple de détecteurs placés dans la maison, cela revient à vouloir minimiser le nombre de détecteurs nécessaires et donc de faire des économies sur l'achat des détecteurs. L'exemple de la maison est "simple" dans le sens où la maison ne contient pas beaucoup de pièces. Donc il n'y a pas beaucoup de sommets dans le graphe. Nous pouvons aussi considérer des graphes qui correspondent à des bâtiments plus complexes, comme des châteaux, des musées ou encore des complexes hôteliers. Plus le nombre de pièces augmentent, plus l'intérêt de minimiser le nombre de détecteurs nécessaires est évident.

Le problème de trouver un code identifiant est équivalent à un problème de couverture dans les graphes. Si nous appelons l'attribut d'un sommet donné (c'est-à-dire l'ensemble des sommets voisins) un *disque* dont le *centre* est le sommet donné, alors un code identifiant est le placement de disques tel que chaque sommet est couvert par au moins un disque (cela correspond à la condition de pouvoir détecter tout incendie dans la maison) et tel que l'ensemble des disques qui couvrent un sommet est unique (cela correspond à la condition de pouvoir déterminer la position exacte de l'incendie). Lorsque nous considérons l'ensemble des sommets voisins, nous étudions des disques de rayon 1 (les sommets voisins sont proches du sommet donné). Un autre problème de couverture, auquel nous nous sommes intéressés, traite de disques de rayon r fixé, c'est-à-dire de l'ensemble des sommets qui se trouvent à distance au plus r du sommet donné.

Par exemple, considérons un problème de télécommunication (Figure 5). Dans un système d'antennes téléphoniques pour téléphones portables, nous voulons que deux antennes se trouvent à distance au moins $r+1$ l'une de l'autre (pour éviter les interférences) et nous voulons que chaque téléphone portable soit à distance au plus r de deux antennes différentes (afin de garantir une bonne qualité de transmission). En terme de graphes, chaque antenne est le sommet d'un centre d'un disque de rayon r et tout sommet qui ne correspond pas à un centre de disque doit être recouvert par deux disques. Cet exemple est un problème de couverture appelé (r, a, b)-code couvrant où a désigne le nombre d'antennes qui se trouvent dans chaque disque et b désigne le nombre de disques qui recouvrent chaque sommet correspondant à un téléphone portable. Ces codes ont été introduits par Cohen et al. [CHLM95] sous le nom de *codes couvrants pondérés*.

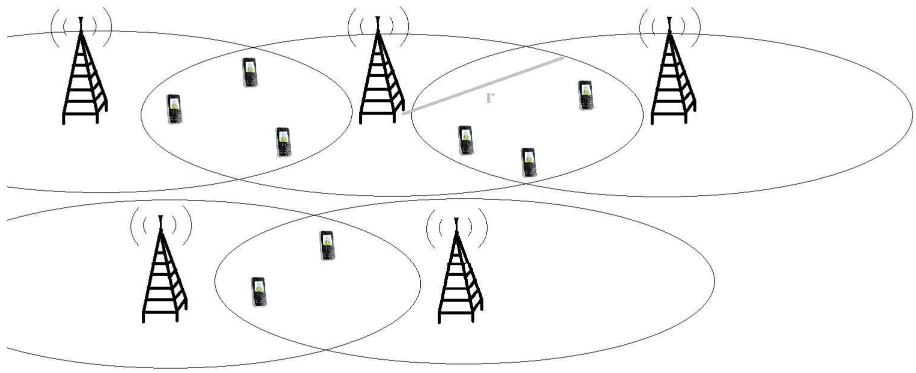


Figure 5: Le placement des antennes par rapport aux téléphones portables correspond à un $(r, 1, 2)$-code couvrant.

Revenons au problème des codes identifiants. Nous avons montré qu'il était équivalent à un problème de couverture dans les graphes. Nous allons maintenant le traduire en un problème de combinatoire des mots dans le cas particulier des graphes appelés chemins P_n de longueur n (Figure 6). Considérons les n sommets $u_1, \ldots, u_n$ du chemin P_n et un code identifiant C de ce chemin. Si le sommet u_i fait partie du code, nous le remplaçons par `1`, sinon par `0`. Le mot obtenu (c'est-à-dire la suite des lettres obtenue) à partir de $u_1 \cdots u_n$ est alors un mot sur l'alphabet $\{\mathtt{0}, \mathtt{1}\}$ et correspond au code identifiant C. Par exemple, le mot `101` correspond à un code identifiant de P_3 où u_1 et u_3 sont les sommets du code.

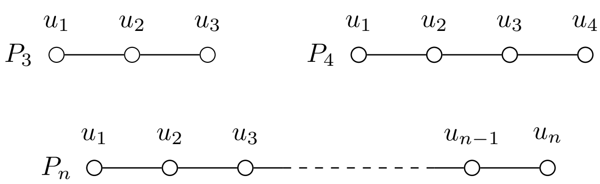


Figure 6: Les graphes P_3, P_4 et P_n sont appelés chemins de longueur respective 3, 4 et n.

Considérons à présent le graphe représenté à la Figure 7, où des flèches étiquetées relient les sommets. Un cheminement (ou parcours) dans ce graphe est l'enchaînement de plusieurs flèches en respectant le sens des flèches. Comptons le nombre de chemins empruntant n arcs, commençant par le sommet "Début" et se terminant dans un des sommets grisés. Nous pouvons montrer que ce nombre est exactement le nombre d'ensembles de sommets formant un code identifiant dans le chemin P_n. De plus, si nous considérons les étiquettes des arcs empruntées par un de ces cheminements particuliers, il s'agit du mot correspondant à un code identifiant d'un chemin P_n. Nous remarquons que tous ces mots commencent par 111, 101 ou 011 et se terminent par 111, 110 ou 101. Dans ces mots, nous pouvons nous intéresser aux occurrences de certains mots finis. Par exemple, ces mots ne contiennent jamais le mot 000 (car il n'y a pas trois flèches consécutives qui sont étiquetées par un 0).

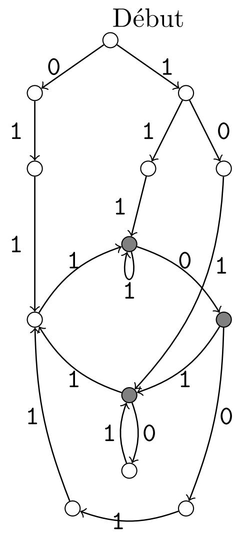


Figure 7: Des flèches étiquetées relient les sommets du graphe représentant les codes identifiants dans le chemin P_n.

Dans le cas général des mots infinis, nous pouvons aussi compter le nombre de mots finis apparaissant dans un mot infini donné. Nous considérons des mots infinis obtenus en appliquant un nombre infini de fois certaines règles. Une règle possible est de remplacer chaque 0 par 01 et chaque 1 par 10. Par exemple, si nous appliquons cette règle 3 fois à partir du mot 0, nous obtenons

$$\mathtt{0} \to \mathtt{01} \to (\mathtt{01})(\mathtt{10}) = \mathtt{0110} \to (\mathtt{01})(\mathtt{10})(\mathtt{10})(\mathtt{01}) = \mathtt{01101001}.$$

Cette règle permet de construire un mot infini très célèbre, appelé le mot de Thue–Morse. Dans ce mot, nous comptons le nombre de mots finis d'une longueur donnée qui apparaissent. Ces mots sont appelés *facteurs* du mot infini. Par exemple, 00, 01, 10 et 11 sont les 4 facteurs de longueur 2 du mot infini de Thue–Morse. Ces mots apparaissent déjà après la quatrième itération de la règle. Au lieu de compter les facteurs distincts, nous pouvons compter les facteurs distincts à une équivalence près. Considèrons l'équivalence “être un anagramme l'un de l'autre”, qui porte le nom d'*équivalence abélienne*. Dans le mot the Thue–Morse, il n'y en a que trois facteurs distincts de longueur 2 lorsqu'ils sont comptà équivalence abélienne près. En effet, 01 est l'anagramme de 10. Une généralisation de l'équivalence abélienne a été introduite par Karhumäki et al. [KSZ13] sous le nom d'équivalence ℓ-abélienne. Un des problèmes étudiés dans cette thèse traite du nombre de facteurs apparaissant dans le mot de Thue–Morse, lorsqu'ils sont comptés à l'équivalence 2-abélienne près.

Le dernier problème concerne la version abélienne de la notion de mot de retour dans les mots infinis. Soit u un facteur d'un mot infini donné. A mot de retour à u est un facteur qui commence à une occurrence de u et qui se termine juste avant l'occurrence suivante de u dans le mot infini. Par exemple, si nous considérons le mot de Thue–Morse et $u = 011$, alors nous notons par des barres les occurrrences de u dans le début du mot de Thue–Morse comme suit

$$|011010|011001|01101001|0110|011010|011001|0110|01101001|011010|011.$$

Le premier mot de retour à u est alors le mot 011010, le second mot de retour à u est 011001 etc. Comme dans le problème précédent, nous sommes intéressé par cette notion dans un contexte abélien. Nous considérons donc les occurrences de u à équivalence abeliènne près. Cette question a déjà été étudiée par Puzynina et Zamboni [PZ13].

Résumé

Cette thèse se divise en deux parties qui sont le reflet de la cotutelle. La première partie se consacre à des questions de combinatoire des mots dans un contexte abélien, tandis que la seconde se préoccupe de problèmes de couvertures dans des graphes. Chaque partie est organisée de la même façon et commence par un chapitre reprenant les notions de base. Ce premier chapitre donne aussi une première introduction aux problèmes qui nous intéressent. Chaque problème particulier, accompagné de son contexte, fait l'objet d'un chapitre. Ces chapitres se terminent par une section contenant des questions ouvertes ou des pistes de recherche. Après les deux parties se trouvent les annexes qui contiennent des compléments d'informations et de démonstrations.

Chapitre 1 : combinatoire des mots

Nous supposons le lecteur familier avec les notions de base de combinatoire des mots. Les travaux de cette thèse se placent dans un contexte abélien. Le but de ce chapitre est de rappeler les notions d'équivalence ℓ-abélienne, de k-régularité et de mots sturmiens.

Deux mots finis sont abéliens équivalents si l'un est l'anagramme de l'autre. Cette notion d'équivalence abélienne a été généralisée par Karhumäki et al. [KSZ13]. Soit ℓ un entier strictement positif. Deux mots finis u et v sont ℓ-abéliens équivalents si pour tout facteur x de longueur au plus ℓ, le nombre d'occurrences de x dans u coïncide avec le nombre d'occurrences de x dans v. Cette définition implique que l'équivalence 1-abélienne est exactement l'équivalence abélienne traditionnelle.

Une notion, utilisée dans le chapitre suivant, est celle de la régularité d'une suite d'entiers au sens défini par Allouche et Shallit [AS03b]. Soit k un entier strictement plus grand que 2. Une suite $\mathbf{s} = (s_n)_{n\geq 0}$ d'entiers est k-régulière s'il existe un nombre fini de suite $t_1(n)_{n\geq 0}, \ldots, t_\ell(n)_{n\geq 0}$ tel que toute suite appartenant à son k-noyau

$$\mathcal{K}_k(\mathbf{s}) = \{s(k^e n + r)_{n\geq 0} : e \geq 0 \text{ et } 0 \leq r \leq k^e - 1\}$$

peut s'écrire comme une $\mathbb{Z}$-combinaison linéaire des $t_i(n)_{n\geq 0}$. Une méthode pour prouver la k-régularité d'une suite $\mathbf{s} = (s_n)_{n\geq 0}$ est de montrer que toutes les sous-suites $s(k^e n + r)_{n\geq 0}$ pour un e fixé et $0 \leq r \leq k^e - 1$ s'expriment comme combinaison linéaire de sous-suites de la forme $s(k^{e'} n + r')_{n\geq 0}$ avec $e' < e$ et $0 \leq r' \leq k^{e'} - 1$.

En combinatoire des mots, les mots sturmiens forment une des classes les plus étudiées des mots infinis [Lot02, PF02]. Il s'agit des mots infinis ayant la plus petite complexité en nombre de facteurs parmi tous les mots apériodiques. Un des exemples les plus connus de mots sturmiens est le mot de Fibonacci $\mathbf{f} = \mathtt{01001010010010100}\cdots$. Nous rappelons plusieurs caractérisations des mots sturmiens en termes de mots apériodiques et équilibrés (Théorème de Coven–Hedlund), de complexité ℓ-abélienne [KSZ13], de mots de retour [Vui01]. Nous terminons par la caractérisation des mots sturmiens en tant que codages de l'orbite d'un point, appelé intercepte, sur le cercle unitaire sous l'action d'une rotation d'angle irrationnel.

Chapitre 2 : régularité des complexités ℓ-abéliennes

Dans ce chapitre, nous nous intéressons aux suites d'entiers et à leur structure. Il y a un an, Madill et Rampersad se sont intéressés à la complexité abélienne du mot de pliage de papier ordinaire [MR13]. Ils ont montré que celle-ci était 2-régulière en exhibant un nombre fini de relations de récurrence. Les auteurs ont ainsi été les premiers à calculer précisément une complexité abélienne qui n'est pas bornée. Depuis, d'autres complexités abéliennes (non-bornées) ont été étudiées. Karhumäki et al. ont étudié le comportement asymptotique de la complexité ℓ-abélienne du mot de Thue–Morse et ont aussi exhibé des relations de récurrence satisfaites par la complexité abélienne du mot "period-doubling" $\mathbf{p}$ [KSZ]. Ces relations montrent que la complexité abélienne de $\mathbf{p}$ est 2-régulière. Dans [BSCRF14], la complexité abélienne d'un point fixe $\mathbf{v}$ du morphisme non-uniforme $\mathtt{0} \mapsto \mathtt{012}, \mathtt{1} \mapsto \mathtt{02}, \mathtt{2} \mapsto \mathtt{1}$ est étudiée et les auteurs obtiennent des résultats similaires à ceux développés dans ce chapitre.

Nous montrons d'abord que toute suite qui satisfait une certaine relation de récurrence avec un paramètre c et 2^{ℓ_0} conditions initiales est 2-régulière (Théorème 2.11). En particulier, ces suites satisfont une certaine propriété de symétrie sur chaque intervalle $[2^\ell, 2^{\ell+1})$ pour $\ell \geq \ell_0$. Ils semblent que plusieurs fonctions de complexité abélienne et 2-abélienne satisfont une telle symétrie.

Ensuite, nous utilisons ce résultat pour montrer que le mot de Thue–Morse possède une complexité 2-abélienne qui est 2-régulière. Pour cela, nous commençons par étudier la complexité abélienne du mot $\mathbf{y}$ qui est le codage par blocs de longueur 2 du mot de Thue–Morse. Cette complexité est étroitement liée à la complexité 2-abélienne du mot de Thue–Morse. Elle satisfait une relation de récurrence similaire à celle donnée dans le Théorème 2.11 (Proposition 2.28). Nous introduisons la quantité $\Delta_{12}(n)$ qui est la différence entre le nombre maximal de lettres $\mathtt{1}$ et $\mathtt{2}$ et le nombre minimal de ces lettres apparaissant dans les facteurs de longueur n de $\mathbf{y}$. Nous montrons que cette suite $\Delta_{12}(n)_{n\geq 0}$ satisfait une récurrence de la même forme que celle énoncée dans le Théorème 2.11 et est donc 2-régulière (Proposition 2.20 et Corollaire 2.21). De plus, il apparaît que $\Delta_{12}(n) + 1 = \mathcal{P}^{(1)}_{\mathbf{p}}$. Nos résultats peuvent donc être reliés aux travaux développés dans [BSCRF14] et [KSZ]. La régularité de la complexité abélienne $\mathcal{P}^{(1)}_{\mathbf{y}}$ de $\mathbf{y}$ découle de celle de Δ_{12}. De même, la régularité de la complexité 2-abélienne du mot de Thue–Morse provient de la régularité de suites $\mathcal{P}^{(1)}_{\mathbf{y}}$ et Δ_{12}.

Des arguments similaires nous permettent de montrer que la complexité 2-abélienne du mot "period-doubling" est elle aussi 2-régulière. Cette similarité suggère l'existence d'un cadre général. Dans le cas du mot "period-doubling", la quantité adéquate à introduire est $\Delta_0(n)$ qui est la différence entre le nombre maximal de lettres $\mathtt{0}$ et le nombre minimal de lettres $\mathtt{0}$ apparaissant dans les facteurs de longueur n du mot $\mathbf{x}$ qui est le codage par blocs de longueur 2 du mot "period-doubling" $\mathbf{p}$. Nous montrons que les suites $\Delta_0(n)_{n\geq 0}$ et $\mathcal{P}^{(1)}_{\mathbf{p}}$ satisfont chacune une relation de récurrence de la forme donnée dans le Théorème 2.11.

Par conséquent, elles sont toutes deux 2-régulières. Comme dans le cas de Thue–Morse, la régularité de la complexité 2-abélienne du mot "period-doubling" découle de la régularité des suites $\mathcal{P}_{\mathbf{x}}^{(1)}$ et Δ_0.

Vu la similarité des arguments, nous pouvons espérer qu'il existe un cadre plus général qui nous permettraient de traiter en une fois ces problémes de régularité des suites de complexités. D'ailleurs, quelques calculs par ordinateur semblent indiquer que l'approche présentée permettrait de montrer que la complexité 3-abélienne du mot de Thue–Morse ainsi que celle du mot "period-doubling" satisfont des relations de récurrence ainsi qu'une propriété de réflexion sur les intervalles $[2^\ell + 2, 2^{\ell+1} + 2]$.

Chapitre 3 : mots de retour abélien

Les mots de retour sont des notions classiques en combinatoire des mots ainsi que dans les systèmes symboliques dynamiques [Dur98, HZ99, JV00, Vui01]. Pour un facteur u d'un mot infini $\mathbf{x}$, un mot de retour à u est un facteur w qui commence par u et se termine juste avant l'occurrence suivante de u. Si $\mathbf{x}$ est uniformément récurrent, c'est-à-dire si chaque facteur de $\mathbf{x}$ apparaît infiniment souvent et à distance bornée, l'ensemble des mots de retour à u est fini. Sous cette hypothèse, l'ensemble des mots de retour forme un code [Dur98] et le mot infini $\mathbf{x}$ peut s'écrire de manière unique comme une concaténation de mots de retour $\mathbf{x} = m_0 m_1 m_2 \cdots$. Nous pouvons ordonner ces mots dans l'ordre de leur première occurrence dans $\mathbf{x}$ et ainsi définir une application $\Lambda_{\mathbf{x},u}$ qui à un mot de retour associe son ordre. La suite dérivée $\mathcal{D}_u(\mathbf{x})$ est alors le mot infini $\Lambda_{\mathbf{x},u}(m_0)\Lambda_{\mathbf{x},u}(m_1)\Lambda_{\mathbf{x},u}(m_2)\cdots$ sur l'alphabet $\{1, \ldots, \#\mathcal{R}_{\mathbf{x},u}\}$. Cette notion de suites dérivées a été utilisée notamment par Durand pour caractériser les mots primitifs substitutifs (Théorème 1.58). Les mots de retour ont aussi été utilisés par Vuillon [Vui01] pour caractériser les mots sturmiens (Théorème 1.59) ou encore les mots périodiques (Proposition 1.60).

Dans ce chapitre, nous considérons la version abélienne des mots de retour. Cette notion a déjà été étudiée par Puzynina et Zamboni [PZ13] et présentée lors de la conférence WORDS 2011. Leur résultat principal est une caractérisation des mots sturmiens (Théorème 3.1). Les auteurs s'intéressent aussi au lien entre le nombre de mots de retour abélien et la périodicité. Ils fournissent une condition suffisante pour la périodicité (Lemme 3.2).

La différence principale entre [PZ13] et le travail présenté dans ce chapitre est que nous considérons l'ensemble des mots de retour par rapport à tous les facteurs d'un mot infini, tandis que Puzynina et Zamboni étudient l'ensemble des mots de retour par rapport à chaque facteur pris séparément.

Nous nous intéressons aussi au lien entre périodicité et mots de retour abélien. Nous considérons tout d'abord l'ensemble $\mathcal{AR}_{\mathbf{x}}$ des mots de retour abélien par rapport à tous les facteurs du mot donné. Nous montrons qu'un mot récurrent est périodique si et seulement si $\mathcal{AR}_{\mathbf{x}}$ est fini (Théorème 3.7). La récurrence uniforme implique bien entendu la récurrence abélienne uniforme, mais le contraire n'est pas vrai en général. Nous donnons un exemple d'un mot qui est récurrent abéliennement uniformément, mais qui n'est pas uniformément récurrent (Proposition 3.8). Notons que ce mot est construit à partir du mot de Thue–Morse.

Ensuite, nous considérons l'ensemble $\mathcal{APR}_{\mathbf{x}}$ des mots de retour abélien par rapport à tous les préfixes du mot donné. Contrairement au caractère fini de l'ensemble $\mathcal{AR}_{\mathbf{x}}$, le caractère fini de $\mathcal{APR}_{\mathbf{x}}$ n'implique pas la périodicité ni la périodicité abélienne de $\mathbf{x}$. Nous étudions le cas particulier du mot de Thue–Morse $\mathbf{t}$ et montrons que $\mathcal{APR}_{\mathbf{t}}$ contient 16 éléments

(Théorème 3.15). Nous obtenons une caractérisation des mots sturmiens avec un intercepte non-nul en termes du caractère fini de l'ensemble des mots de retour abélien par rapport aux préfixes (Theorem 3.18). Comme le mot de Fibonacci $\mathbf{f}$ est un mot sturmien avec une pente et un intercepte égaux à $1/\phi^2$, où ϕ est le nombre d'or, notre résultat implique que l'ensemble $\mathcal{APR}_{\mathbf{f}}$ est fini. Nous décrivons explicitement cet ensemble qui contient cinq éléments. En comparaison, l'ensemble des mots de retour abélien par rapport aux préfixes du mot $\mathtt{0}\mathbf{f}$ est infini. Nous terminons notre étude de l'ensemble $\mathcal{APR}_{\mathbf{x}}$ en montrant que si $\mathbf{x}$ est un mot récurrent abéliennement tel que $\mathcal{APR}_{\mathbf{x}}$ est fini, alors la complexité abélienne de $\mathbf{x}$ est finie.

Enfin, nous considérons l'analogue abélien de la suite dérivée. Si $\mathbf{x}$ est uniformément récurrent, il peut alors être factorisé en termes des mots de retour abélien par rapport à un préfixe donné de $\mathbf{x}$. Cela induit un codage qui donne lieu à une nouvelle suite. Contrairement au cas non-abélien, la factorisation de $\mathbf{x}$ en mots de retour abélien n'est pas nécessairement unique (Exemple 3.27). Une autre différence entre le cas "classique" et le cas abélien est que la caractérisation des mots primitifs substitutifs obtenue en terme de la suite dérivée par Durand n'est plus valable dans le cas abélien. Le mot de Thue–Morse est un exemple de mot primitif substitutif ayant infiniment de suites dérivées abéliennes (Proposition 3.36).

Chapitre 4 : théorie des graphes

Nous supposons que le lecteur est familier avec les notions de base de théorie des graphes. Le but de ce chapitre est de rappeler deux problèmes de couverture dans les graphes.

Les codes identifiants ont été introduits en 1998 par Karpovsky, Chakrabarty et Levitin [KCL98] pour modéliser un problème pratique d'identification de processeurs défectueux dans des réseaux multiprocesseurs. Plus spécifiquement, le réseau multiprocesseur est représenté par un graphe dont les sommets sont les processeurs. Imaginons que chaque processeur est capable de tester si un processeur dans son voisinage fermé (i.e., les processeurs voisins et lui-même) est défectueux et ne puisse retourner qu'une information binaire. Par exemple, un processeur renvoie 0 si aucune défaillance n'a été détectée et 1 dans les autres cas. Le problème est de déterminer un sous-ensemble C de processeurs tel que

- si tous les processeurs de C renvoient l'information 0, cela signifie qu'il n'y a pas de défaillance,
- si au moins un des processeurs de C renvoie l'information 1, il y a une défaillance et nous pouvons localiser le processeur défectueux de manière unique.

En supposant qu'à tout moment, il y a au plus un processeur défectueux, l'ensemble C recherché correspond exactement à un code identifiant.

En effet, la première condition garantit que s'il y a une défaillance, elle sera détectée. En termes de graphes, cela signifie que C est un ensemble dominant, c'est-à-dire que tout sommet du graphe est dans C ou est voisin d'un sommet de C. La deuxième condition, qui permet de localiser de manière unique le processeur défectueux, équivaut à dire que C est un ensemble séparant du graphe.

Ils existent aussi d'autres applications des codes identifiants. Par exemple, ces codes sont utilisés pour modéliser un problème de localisation par des réseaux de capteurs [RSTU04] ou pour concevoir des réseaux de détecteurs d'incendie dans des bâtiments [UTS04].

Enfin, les codes identifiants dans des graphes fortement réguliers sont étroitement liés aux ensembles résolvants introduits par [Bab80]. Un ensemble résolvant est un ensemble de sommets S tel que chaque sommet du graphe est déterminé de manière unique par ses distances aux sommets de S. La taille minimale d'un tel ensemble est appelé dimension métrique. Dans le cas de graphes avec un diamètre égal à 2, un ensemble résolvant C est presque un code identifiant : seul les sommets du code ne sont éventuellement pas identifiés. Auquel cas, il suffit d'ajouter $|C|$ sommets à l'ensemble pour le rendre identifiant.

Le second problème de couverture auquel nous nous intéressons traite de couverture par des boules de rayon r. Pour des entiers positifs r, a et b, un (r, a, b)-code couvrant d'un graphe $G = (V, E)$ est un ensemble S de sommets qui correspondent aux centres des boules de la couverture et tel que la propriété suivante est satisfaite. Tout sommet de S (respectivement de $V \setminus S$) est recouvert par a (resp. b) boules de rayon r. Ces codes sont aussi connus sous le nom de (r, a, b)-coloriages isotropes [Axe03] ou coloriages parfaits [Puz08]. Ses codes couvrants généralisent la notion de code parfait. En effet, un r-code est un (r, a, b)-code avec $a = 1 = b$. Golomb et Welch [GW68, GW70] ont démontré l'existence de $(1, 1, 1)$-code couvrant dans la grille multi-dimensionnelle $\mathbb{Z}^d$. De plus, les auteurs conjecturent qu'il n'existe pas de $(r, 1, 1)$-codes couvrants dans $\mathbb{Z}^d$.

Les (r, a, b)-codes couvrants ont déjà été étudiés pour le nom de codes couvrants pondérés par Cohen et al. [CHLM95]. Les auteurs considèrent ces codes dans la métrique de Hamming, mais les (r, a, b)-codes couvrants ont été aussi étudiés dans des graphes en général [Tel94]. Telle [Tel94] donne notamment un résultat de complexité: le problème de décision à propos de l'existence de $(1, a, b)$-code dans un graphe est NP-complet. Le cas particulier des (r, a, b)-codes couvrants avec $r = 1$ a fait l'objet de beaucoup de recherches. Par exemple, Dorbec et al. [DGHM09] et Gravier et al. [GMP99] ont considérés les $(1, a, b)$-codes couvrants dans le cas de la métrique de Lee. Dorbec et al. [DGHM09] présentent notamment une construction pour obtenir des $(1, a, b)$-codes couvrants dans la grille multi-dimensionnelle $\mathbb{Z}^d$. Dans le cas particulier de la grille 2-dimensionnelle $\mathbb{Z}^2$, Puzynina [Puz04, Puz08] a obtenu des résultats de périodicité pour les (r, a, b)-codes couvrants et Axenovich [Axe03] a donné une caractérisation de ces codes lorsque $r > 2$ et $|a - b| > 4$.

Chapitre 5 : codes identifiants dans des graphes sommets-transitifs

Le problème de trouver un code identifiant de taille minimale est NP-complet en général [CHL03] mais peut être exprimé comme un programme linéaire entier (Remarque 4.26). Nous considérons la relaxation fractionnaire de ce programme. Cela revient à attribuer des poids réels compris entre 0 et 1 à chaque sommet, au lieu d'attribuer un poids 1 à chaque sommet dans le code et 0 aux autres. Naturellement, nous sommes intéressés par la "taille" minimale d'un code identifiant fractionnaire, c'est-à-dire par minimiser la somme de tous les poids.

Il est naturel de vouloir comparer la taille minimale fractionnaire γ_f^{ID} avec la taille minimale entière γ^{ID}. Leur rapport vaut au plus un facteur logarithmique en le nombre de sommets du graphe (voir Proposition 5.1).

Dans le cas des graphes sommets-transitifs, i.e., des graphes où tous les sommets jouent le même rôle, nous pouvons calculer la taille minimale. Elle dépend de seulement trois paramètres du graphe : le nombre de sommets, le degré des sommets et la plus petite taille des différences symétriques de deux voisinages fermés. Les codes identifiants entiers ont déjà été beaucoup étudiés dans les graphes sommets-transitifs, particulièrement dans les cycles [BCHL04, GMS06, JL12, XTH08] et dans les hypercubes [BHL00, EJLR08, ELR08, HL02, KCL98]. Dans ces exemples, l'ordre de la taille minimale entière γ^{ID} semble toujours être égal à l'ordre de la taille minimale fractionnaire. Cependant, la plus petite taille des différences symétriques de voisinages fermés est petite par rapport au nombre de sommets : soit elle est constante (pour les cycles), soit elle est d'ordre logarithmique dans le nombre de sommets (pour les hypercubes). Pour cette raison, nous nous concentrons sur les graphes fortement réguliers et sommets-transitifs. Ces graphes ont comme propriété que la taille des différences symétriques est toujours au moins d'ordre $\sqrt{|V|}$ (si le graphe n'est pas trivial).

De plus, tout graphe fortement régulier est de diamètre 2. Par conséquent, dans ces graphes, la taille minimale entière γ^{ID} est toujours du même ordre que la dimension métrique. Nous pouvons alors calculer l'ordre de la taille minimale entière dans les graphes fortement réguliers dont la dimension métrique est connue. C'est le cas pour les graphes de Kneser, de Johnson (de diamètre 2) et pour les graphes de Paley. Ces familles de graphes exhibent deux comportements différents par rapport aux codes identifiants. Les graphes de Kneser et de Johnson de diamètre 2 sont une famille de graphes avec γ^{ID} et γ_f^{ID} qui sont du même ordre $\sqrt{|V|}$, tandis que les graphes de Paley ont une taille minimale fractionnaire γ_f^{ID} bornée par une constante et une taille minimale entière γ^{ID} d'ordre $\log_2(|V|)$. En d'autres termes, pour les graphes de Paley, γ^{ID} et γ_f^{ID} sont éloignés d'un facteur logarithmique.

Dans la section 5.4, nous considérons une famille de graphes qui n'a encore jamais été étudiée de manière générale dans le cadre des codes identifiants ou des ensembles résolvants. Il s'agit des graphes d'adjacence de quadrangles généralisés. Un quadrangle généralisé est une structure d'incidence, c'est-à-dire un ensemble de points et de lignes, telle que chaque ligne contient un nombre fixe de points, chaque point appartient à un nombre fixe de lignes et tout point se projette de manière unique sur une ligne à laquelle il n'appartient pas. Par exemple, un graphe complet biparti est un graphe d'adjacence d'un quadrangle généralisé, le produit cartésien de deux cliques de même taille en est un aussi. Dans le dernier exemple, Gravier et al. [GMS08] ont déjà étudié les codes identifiants et les deux valeurs γ^{ID} et γ_f^{ID} sont de même ordre. Lorsque le graphe d'adjacence d'un quadrangle généralisé est sommet-transitif, nous établissons des bornes sur la taille minimale fractionnaire γ_f^{ID} (voir Proposition 5.12).

A l'heure actuelle, les constructions de quadrangles généralisés sont connues pour seulement certaines valeurs des paramètres correspondants au nombre de points sur chaque ligne et au nombre de lignes passant par chaque point. Ces constructions font appel à la géométrie finie. Nous construisons des codes identifiants d'ordre optimal dans certains quadrangles généralisés[1]. Ces ordres sont de la forme $|V|^{\alpha}$ avec $\alpha \in \{\frac{1}{4}, \frac{1}{3}, \frac{2}{5}\}$ et correspondent à l'ordre de taille minimale fractionnaire.

En conclusion, nous avons considéré des familles de graphes qui ont soit des valeurs γ^{ID} et γ_f^{ID} du même ordre, soit des valeurs γ^{ID} et γ_f^{ID} éloignées d'un facteur logarithmique. Cependant, dans le dernier cas, la taille minimale fractionnaire γ_f^{ID} est bornée par une constante. Il serait intéressant de déterminer s'il existe une famille infinie de graphes telles que le rapport entre γ^{ID} et γ_f^{ID} serait logarithmique et que γ_f^{ID} ne serait pas borné par une constante. Une question similaire est de déterminer s'il existe un graphe (ou une famille de graphes) tel que le rapport de la solution entière et de la solution fractionnaire ne soit ni constant, ni logarithmique. Enfin, comme les graphes considérés sont de diamètre 2, nos résultats peuvent être étendus aux ensembles localisateurs-dominateurs et à la dimension métrique.

[1] Toutes les constructions de graphes considérées donnent lieu à des graphes d'adjacence sommets-transitifs.

Chapitre 6 : 2-coloriages constants et une application aux (r, a, b)-codes couvrants de la grille infinie 2-dimensionnelle

Dans ce chapitre, nous introduisons la notion de 2-coloriage constant. Il s'agit de coloriages avec deux couleurs, noir et blanc, de graphes pondérés qui satisfont certaines propriétés. Pour un sous-groupe A d'automorphismes d'un graphe pondéré G et un sommet fixé v de G, un 2-coloriage c est constant si, pour tout coloriage $c \circ f$ avec $f \in A$, la somme des poids des sommets noirs ne prend qu'au plus deux valeurs :

- la valeur a si v est de couleur noire,
- la valeur b si v est de couleur blanche.

Une question naturelle est "étant donné un graphe pondéré, un sommet particulier et un sous-groupe d'automorphismes, existe-t-il un 2-coloriage constant ?". Nous considérons principalement des cycles pondérés avec au plus 3 poids différents et nous fixons le sous-groupe d'automorphismes à l'ensemble des rotations du cycle. Pour quatre types de cycles, nous caractérisons les constantes a et b des 2-coloriages constants possibles, en fonction des poids des sommets.

La motivation pour définir les 2-coloriages constants provient des problèmes couvrants dans les graphes et plus particulièrement des (r, a, b)-codes couvrants dans la grille infinie multidimensionnelle. Pour un rayon égal à 1, Dorbec et al. [DGHM09] présentent une méthode pour construire des $(1, a, b)$-codes couvrants. Cette méthode produit des codes périodiques. Dans le cas particulier de la grille infinie 2-dimensionnelle, Puzynina [Puz04] a montré qu'il existe des $(1, a, b)$-codes couvrants qui ne sont pas périodiques, mais que pour ces valeurs de a et b, il existe des coloriages périodiques qui sont des $(1, a, b)$-codes couvrants. Toujours dans la grille 2-dimensionnelle, Puzynina [Puz08] a considéré des (r, a, b)-codes couvrants avec un rayon plus grand ou égal à 2. L'auteur a démontré que tout ces (r, a, b)-codes couvrants sont périodiques [Puz08]. De plus, Axenovich [Axe03] a donné une caractérisation des (r, a, b)-codes couvrants de $\mathbb{Z}^2$ avec $r \geq 2$ et $|a - b| > 4$ en termes de coloriages diagonaux. La définition des 2-coloriages constants est apparue naturellement pour traduire cette périodicité. Nous nous concentrons dans ce chapitre sur la grille infinie 2-dimensionnelle afin d'utiliser la caractérisation obtenue par Axenovich [Axe03]. Dans cette caractérisation, les valeurs exactes de a et b ne sont pas données. Notre approche avec les 2-coloriages constants permet de combler cette lacune.

Nous montrons que les (r, a, b)-codes couvrants de la grille infinie sont étroitement liés aux 2-coloriages constants de cycles pondérés. Pour ce faire, nous présentons une méthode de projection et repliement qui nous permet de réduire la donnée de la grille infinie en un cycle pondéré. En composant la caractérisation obtenue par Axenovich avec cette méthode, les cycles pondérés font partie des quatre types de cycles étudiés. Il ne reste plus qu'à appliquer les résultats obtenus dans le cadre des 2-coloriages constants pour obtenir les valeurs de a et b possibles.

Nous avons donc traduits la périodicité des (r, a, b)-codes couvrants de la grille infinie en 2-coloriages constants de quatre types de cycles pondérés. Il serait intéressant d'essayer de traduire la périodicité des $(1, a, b)$-codes couvrants de la grille en d dimensions en termes de 2-coloriages constants. Une autre piste de recherche serait de remplacer la grille infinie en deux dimensions par la grille du roi infinie. Dans ce cas, une première étape serait d'obtenir une caractérisation similaire à celle obtenue par Axenovich.

Annexe A : régularité et complexité ℓ-abélienne

Dans cette annexe, nous présentons d'abord le code `Mathematica` utilisé pour conjecturer les relations satisfaites par la complexité 2-abélienne du mot de Thue–Morse. Le calcul de cette complexité est basé sur une approche matricielle ne nécessitant pas de calculer un long préfixe du mot de Thue–Morse. Cette méthode nous permet de calculer les 65538 premiers termes de la complexité 2-abélienne en seulement 4 minutes sur un ordinateur standard (Intel® Core™ i3). En comparaison, si l'on calcule la complexité 2-abélienne en passant en revue tous les facteurs de longueur n, pour n fixé, dans un préfixe suffisamment long du mot de Thue–Morse, il est nécessaire d'itérer au moins 11 fois le morphisme pour construire le préfixe requis et le calcul des 513 premiers termes de la complexité 2-abélienne prend déjà plus de 4 minutes.

Ensuite, nous présentons des fonctions de complexité abélienne qui semblent satisfaire une symétrie dans les valeurs prises sur chaque intervalle de la forme $[2^\ell, 2^{\ell+1}]$ avec un entier positif ℓ suffisamment grand. Nous considérons des mots infinis, sur un alphabet de 3 lettres, qui des sont points fixes de morphismes 2-uniformes. Nous calculons ensuite leurs complexités 1-abélienne et 2-abélienne. Parmi les comportements de ces fonctions de complexité, nous remarquons que plusieurs sont ultimement périodiques ou satisfont une symétrie. Parmi les autres comportements, il serait intéressant de regarder la croissance des fonctions de complexité.

Enfin, nous détaillons les preuves omises dans le chapitre 2. Ces preuves concernent des résultats à propos du mot "period-doubling" ou de son codage par blocs de longueur 2 et sont similaires à celles développées dans le cas du mot de Thue–Morse.

Annexe B : 2-coloriages constants dans les cycles pondérés

Cette annexe contient les preuves omises dans le chapitre 6. Il s'agit de preuves concernant les 2-coloriages constants dans le cas de cycles pondérés.

Conclusion

Durant cette thèse, j'ai pu aborder divers problèmes de la combinatoire des mots et de couverture dans les graphes. En plus des travaux présentés dans les chapitres de ce manuscrit, je me suis aussi intéressée à la complexité syntaxique associée à un ensemble ultimement périodique d'entiers représenté dans un système de numération en base entière. Avec mon co-directeur Rigo, un post-doctorant Rampersad et une autre doctorante Lacroix, nous [LRRV12] avons obtenu des bornes sur le nombre d'éléments du monoïde syntaxique asssocié à un ensemble ultimement périodique d'entiers. Généralisant les travaux de Honkala [Hon86] et développant des techniques différentes de celles employées par Bell et al. [BCFR09], cette étude a permis de mener à des procédures de décision sur le caractère périodique d'ensembles d'entiers donnés par des automates finis.

Le fait d'effectuer mon doctorat en cotutelle m'a permis de voyager, de rencontrer deux communautés de chercheurs lors de différentes conférences et de collaborer avec plusieurs équipes, chacune ayant son sujet de prédilection. J'ai particulièrement apprécié ces collaborations scientifiques comme en témoignent mes neuf co-auteurs de cinq nationalités différentes. Le travail effectué a donné lieu à plusieurs publications (à ce jour, deux articles publiés dans des revues internationales avec comité de lecture [LRRV12, RSV13], un acte de conférence internationale avec comité de lecture [RV11] et trois articles soumis [GPR$^+$, GV, PRRV]).

Overview of the thesis

Introduction

This thesis deals with two (distinct but connected) topics that are combinatorics on words and graph theory. Let us first describe a real-life situation. The game "*Guess Who?*" was very popular when I was a child. This is a game for two players. Each player starts the game with a board that includes cartoon images of 24 people with all the images standing up. The game starts with each player selecting a character. The goal of the game is to be the first one to determine which card one's opponent has selected by asking questions on the physical feature of characters. For instance, one player can ask to the other one whether his character wears glasses. Players ask alternately their questions. Now, we imagine a static version where all the questions of a player are asked in a row. Consider for instance the board depicted in Figure 8 (note that this is a simplified example as the number of characters and features are reduced). This board contains six characters that have eventually a walking stick, a hat, glasses or a moustache. How many questions do we have to ask to be sure to identify the secret character? In other words, how many features do we need to select in order to identify the secret character?

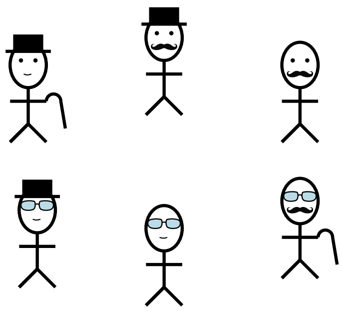


Figure 8: Example of a board fot the game "*Guess Who?*".

In this example, we only need three features, the walking stick, the glasses and the hat, to identify the secret character as shown in Figure 9.

Now we can translate the example in terms of graph theory. Roughly speaking a graph is just a set of points, called vertices, that are linked together by some lines where the points and lines are respectively called vertices and edges. Graphs are useful to model many problems such as the scheme of the underground network of a city, a family tree or a representation of computers network. We can also use them to model the map of a house (Figure 10). The

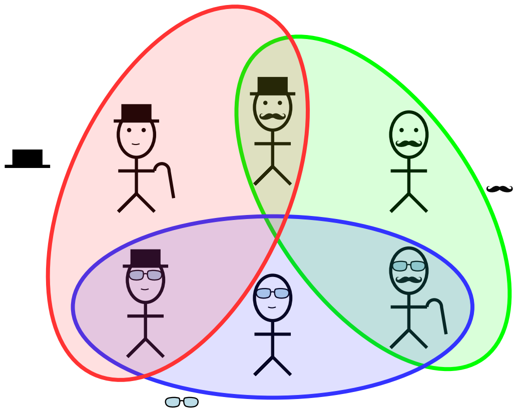


Figure 9: Three features are enough to identify the secret character.

rooms become vertices and two vertices are linked if the corresponding rooms are neighbours of each other.

Figure 10: A house viewed in three different ways: in 3 dimensions, in 2 dimensions and modelled by a graph. The letters L, K and T point the location of the living room, the kitchen and the toilets in each of the views.

The features of a vertex are not physical features as before, but features linked to the structure of the graph. The feature of a vertex is the set of vertices that are linked to the given vertex. Consider again the example of the house. The happy owner of this house wants to protect it from fire issues. So he places fire detectors in some rooms. A detector detects a fire if the latter is in the same room or in a neighbouring room. A natural question is to determine a placement of detectors that allows us to locate precisely the potential fire.

Let us place the detectors in the rooms as prescribed in the picture on the left in Figure 11. These detectors can detect if there is a fire in the house but they can not locate precisely the fire. Indeed, if the detector placed in the living room is the only one to signal a fire, then we do not know if the fire broke out in the kitchen or in the toilets or even in other rooms. To solve this problem, we can add two detectors and place them according to the picture on the right in Figure 11. This placement of detectors still detect any fire and now it allows us to locate precisely the position of the fire. For example, if the detector in the living room is the only one to signal a fire, then we are sure the fire broke out in the toilets. There is no ambiguity anymore.

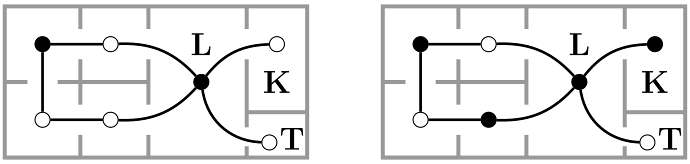


Figure 11: Two placements of fire detectors in the graph that models the house. The black (respectively white) vertices represent the rooms with (resp. withour) a detector.

The set of detectors (i.e., the set of black vertices) forms an *identifying code* of the graph. These codes were introduced in 1998 by Karpovsky, Chakrabarty and Levitin [KCL98] to model fault-diagnosis in multiprocessor systems. Later, other applications of identifying codes were discovered such as the design of emergency sensor networks in facilities [UTS04].

The essential question about identifying codes is to determine the minimum size of such codes in a given graph (i.e., the minimum number of vertices that form an identifying code). In the example of the detectors placed in the house, it means minimizing the number of detectors and so minimizing the cost of detectors. The example presented is "simple" in the sense that the house only has a few rooms. Therefore, its graph only has a few vertices. If we consider more complicated graphs that correspond to more complicated buildings such as a castle, a museum or a hotel complex, then they contain a lot of vertices (rooms) and the interest of minimizing the number of detectors is clear.

We can view the problem of identifying codes as a covering problem. If we call the feature of a given vertex (i.e., the set of neighbouring vertices) a *ball* and the center of this ball is the given vertex, then an identifying code is a placement of balls such that each vertex is covered by at least a ball (it is similar to the condition of detecting any fire in the house) and the set of balls that cover a vertex is unique (it is similar to the condition of locating precisely the place of the fire). When we look at sets of neighbouring vertices, we consider balls of radius 1 (the neighbouring vertices are close to the given vertex). Another covering problem we are interested in considers balls of fixed radius r. Hence vertices that are at some distance r of the given vertex.

For example, let us consider an example of network communication problem (Figure 12). In a system of transmitting stations for cellular phone network, we want each transmitting station to be at distance at least $r + 1$ from each other (to avoid interference) and we want each cellular phone to be within distance r of two distinct transmitting stations (to guarantee a good transmission quality). In terms of graphs, each transmitting station is the center of a ball of radius r and any vertex that does not correspond to a transmitting station must be cover by two balls. This is an example of covering problems called (r, a, b)-covering codes where a is the number of transmitting station in each ball and b is the number of balls that cover any vertex corresponding to a cellular phone. These codes were introduced by Cohen et al. [CHLM95] under the names of *weighted covering codes*.

Consider again the problem of identifying codes. We showed that it was equivalent to a covering problem in graphs. Now we translate it into a problem of combinatorics on words

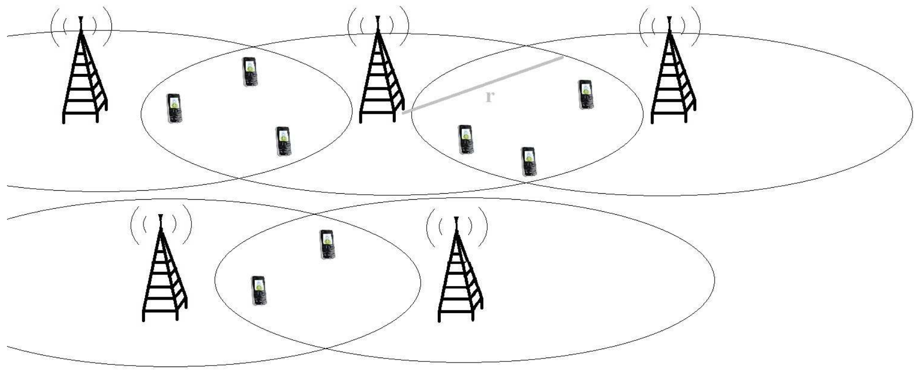


Figure 12: The placement on the transmitting stations with respect to the cellular phones is an example of an $(r, 1, 2)$-covering code.

for the particular case of graphs called paths P_n of length n (Figure 14). Let $u_1, \ldots, u_n$ be the n vertices of the path P_n and let C be an identifying code of P_n. If the vertex u_i belongs to C, then we write 1 instead of u_i. Otherwise, we write 0 instead of u_i. The word (i.e., the sequence of letters) obtained from $u_1 \cdots u_n$ is then a word over the alphabet $\{\mathtt{0}, \mathtt{1}\}$ and it corresponds to the code C. For instance, the word 101 corresponds to an identifying code of P_3 where u_1 and u_3 are vertices of the code.

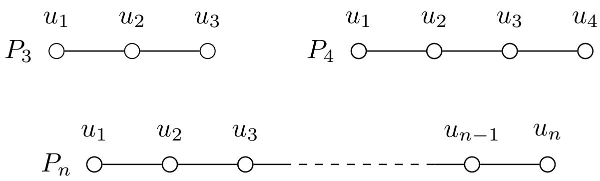


Figure 13: Graphs P_3, P_4 and P_n are called paths of respective length 3, 4 et n.

We look now at the graph depicted in Figure 13 where labelled arrows link the vertices. A route (or journey) through the graph is the chain of consecutive arrows with respect to the direction of the arrows. We count the number of routes going through n arrows, starting with the vertex "Start" and ending in one of the gray vertices. This number is exactly the number of vertices subsets that form identifying codes. Moreover, if we look at the labels of the taken arrows for a particular route, then it is a word corresponding to an identifying code of the path P_n. We observe that any word corresponding to an identifying code starts with 111, 101 or 011 and ends with 110,110 and 101. In theses words, we can observe the number of finite words occurring. For instance, 000 does not occur in any word corresponding to an identifying code (as there do not exist three consecutive arrows with label 0).

In the case of infinite words, we can also consider the number of distinct finite words occurring in a given infinite words. We consider infinite words obtained by applying infinitely many times a rule such as "replace each 0 by 01 and each 1 by 10". For example, if we apply

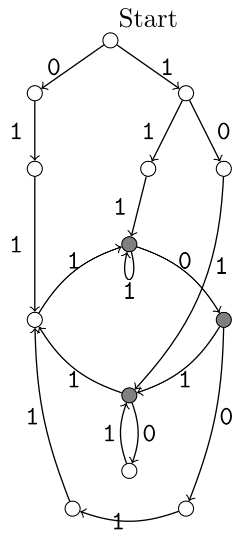


Figure 14: Labelled arrows link the vertices of the graph that represents identifying codes in the path P_n.

three times this rule to the word `0`, we obtain

$$\mathtt{0} \to \mathtt{01} \to (\mathtt{01})(\mathtt{10}) = \mathtt{0110} \to (\mathtt{01})(\mathtt{10})(\mathtt{10})(\mathtt{01}) = \mathtt{01101001}.$$

This rule allows us to construct a well-known infinite word called the *Thue–Morse word*. The finite words occurring in a given infinite word are called *factors* of the infinite word. For instance, `00`, `01`, `10` and `11` are the four factors of length 2 occurring in the Thue–Morse word. These words already appear after 3 iterations of the rule. Instead of counting the number of distinct factors, we can count the number of distinct factors up to a given equivalence. Consider the equivalence "to be anagram of each other" which is called the *abelian equivalence*. There are only three factors of length 2 of the Thue–Morse word, when counted up to abelian equivalence. Indeed, `01` is the anagram of `10`. A generalization of the abelian equivalence was introduced by Karhumäki et al. [KSZ13] under the name of ℓ-abelian equivalence. One of the problem studied in this thesis deals with the number of factors occurring in the Thue–Morse word when they are counted up to 2-abelian equivalence.

The last problem focus on the notion of return words in finite words in an abelian context. Let u be a factor of a given infinite word. A return word to u is a factor that starts with u and ends before the next occurrence of u in the infinite word. For instance, if we consider the Thue–Morse word and $u = 011$, we can mark the occurrences of u in the beginning of the Thue–Morse word as follows

$$|\mathtt{011010}|\mathtt{011001}|\mathtt{01101001}|\mathtt{0110}|\mathtt{011010}|\mathtt{011001}|\mathtt{0110}|\mathtt{01101001}|\mathtt{011010}|\mathtt{011}.$$

So the first return word is `011010`, the second return word is `011001` and so on. As for the previous problem, we are interested in an abelian version of the concept. Hence, we consider the occurrences of u up to abelian equivalence. Such problems have already been studied by Puzynina and Zamboni [PZ13].

Organization of the manuscript

The first three chapters focus on combinatorics on words. The three following chapters deal with problems in graph theory. Both parts follow the same organization. Each one begins with a chapter containing the basic notions as well as a first introduction to the problems considered after. Each particular problem is the topic of a dedicated chapter. Some complement information and proofs are given in the appendices.

We now present the content of each chapter[2].

In Chapter 1, we collect some general results and definitions about words and k-regular sequences (in particular stability properties of the set of k-regular sequences under sum and product [AS03b]) that are needed in the other parts of this thesis. We also present the extension of abelian equivalence of words to ℓ-abelian equivalence that was introduced by Karhumäki, Saarela and Zamboni [KSZ13]. We end this chapter with the definition of Sturmian words as the coding of the orbit of a point on the unit circle under rotation by an irrational angle.

Chapter 2 is dedicated to the study of regularity of ℓ-abelian complexity sequences. We follow the lead from Madill and Rampersad [MR13] who studied the abelian complexity of the paperfolding word and showed it was 2-regular. We first show that any sequence satisfying a particular recurrence relation with a parameter c and 2^{ℓ_0} initial conditions is 2-regular. In particular, these sequences exhibit a reflection symmetry in the values taken over each interval of the form $[2^\ell, 2^{\ell+1})$ for $\ell \geq \ell_0$. Then, we focus on the 2-abelian complexity sequence of the Thue–Morse word. To show that this sequence is 2-regular, we first study the abelian complexity sequence of the 2-block coding of the Thue–Morse word and show that it is 2-regular. To achieve this, we consider a quantity $\Delta_{12}(n)$ that is the difference between the maximal number of letters $\mathtt{1}$ and $\mathtt{2}$ together in a factor of length n of $\mathbf{y}$ and the minimal of this number. The sequence $\Delta_{12}(n)_{n\geq\infty}$ satisfies a recurrence relation that enters the framework of our first theorem. Hence, it is 2-regular. Similar arguments applied to the period-doubling word instead of the Thue–Morse word show that the 2-abelian complexity sequence of the period-doubling word is 2-regular.

The return word is a classical notion in combinatorics on words and symbolic dynamical systems [Dur98, HZ99, JV00, Vui01]. We consider abelian returns, that are return words up to abelian equivalence, in Chapter 3. The notion of abelian returns has already been studied by Puzynina and Zamboni [PZ13]. Their main theorem [PZ13] is a characterization of Sturmian words in terms of the number of abelian returns with respect to each factor of the word. This characterization is similar to the one obtained by Vuillon with the classic return words [Vui01]. The main difference between [PZ13] and our work is that we consider the set of abelian returns with respect to all factors of the word together. We first discuss the relationship between periodicity and finiteness of the set of abelian returns with respect to all factors. Then, we restrict ourselves to the set of abelian returns with respect to all prefixes. For the particular case of the Thue–Morse word, this set contains 16 elements. We obtain a characterization of Sturmian words with a null intercept in terms of the finiteness of the set of abelian returns to all prefixes. We give this set, which contains 5 elements, for the Fibonacci word. Finally, as Durand [Dur98] defined the derived sequence for the classic return words, we define the abelian analogue, the abelian derived sequence. Durand [Dur98] gave a characterization of primitive substitutive in terms of the finiteness of the set of derived

[2] This is a short version of the abstract given in French in “Aperçu de la thèse”.

sequences. Such a characterization does not hold in the abelian case. Indeed, the Thue–Morse word is an example of primitive substitutive word that has infinitely many abelian derived sequences.

Chapter 4 contains the basic notions needed for a clear understanding of the two following chapters. We present identifying codes and (r, a, b)-covering codes that belong to the more general class of covering problems. Identifying codes in a graph correspond to coverings of the graph with balls of radius r (i.e., closed neighbourhood) such that each vertex is covered by at least one ball (domination condition) and the set of balls covering a given vertex is unique (separation condition). These codes were introduced by Karpovsky et al. [KCL98] to model fault-diagnosis in multiprocessor system. The (r, a, b)-codes are coverings that satisfy multiplicity conditions. Each center of balls of radius r is covered a times, whereas each vertex that is not the center of any ball is covered b times. These codes are a particular case of weighted covering codes introduced by Cohen et al. [CHLM95].

In Chapter 5, we consider identifying codes. The problem of computing an identifying code of minimal size is NP-complete in general [CHL03] but can be expressed as an integer linear program. In such program, vertices are given weights. The weight is equal to 1 (respectively 0) if the vertex belongs (resp. do not belong) to the set. Minimizing the size of the identifying code is the same as minimizing the sum of the weights. We consider the fractional relaxation of this program. Hence we allow the weights of vertices to be real between 0 and 1. We are also interested in minimizing the sum of the weights for the fractional relaxation. It is natural to compare the integer solution to the fractional solution. Their ratio is at most logarithmic in the number of vertices of the graphs. We focus on vertex-transitive graphs as we are able to compute precisely the fractional solution for them. Identifying codes have already been studied in different classes of vertex-transitive graphs, especially in cycles [BCHL04, GMS06, JL12, XTH08] and hypercubes [BHL00, EJLR08, ELR08, HL02, KCL98]. In these examples, the integer and fractional solutions seem to have the same order. However, the minimal size of the symmetric differences of neighbourhoods is small compared to the number of vertices. On the opposite, in vertex-transitive strongly regular graphs, the size of symmetric differences is at least $\sqrt{|V|}$ where $|V|$ is the number of vertices. Another advantage to look at strongly regular graphs is that their diameter is equal to 2. So results on identifying codes can be extended to results on metric dimension (and vice versa). For Kneser and Johnson graphs of diameter 2, the integer and fractional solutions are both of order $\sqrt{|V|}$, while for Paley graphs, the fractional solution is bounded by a constant and the integer solution is logarithmic in $|V|$. At the end of the chapter, we consider some adjacency graphs of generalized quadrangles. It is the first time that these graphs are studied in the framework of identifying codes. They form infinite families of vertex-transitive graphs with integer and fractional solutions of order $|V|^\alpha$ with $\alpha \in \{1/4, 1/3, 2/5\}$.

The (r, a, b)-covering codes are another covering problem we considered. In the multidimensional grid $\mathbb{Z}^d$, Dorbec et al. [DGHM09] presented a method to obtain $(1, a, b)$-covering codes. This method is based on a 1-dimensional pattern that is extended by translations to colour $\mathbb{Z}^d$. This method leads thus to periodic $(1, a, b)$-covering codes. In particular, in the 2-dimensional grid $\mathbb{Z}^2$, Puzynina [Puz08] showed that every (r, a, b)-covering code with $r \geq 2$ is periodic and Axenovich [Axe03] gave a characterization of those when $|a - b| > 4$. The notion of constant 2-, introduce in Chapter 6, comes up as a natural manner to translate the periodicity of (r, a, b)-codes. Constant 2-labellings are particular 2-colourings of weighted graphs that satisfy the following property. For every composition of the colouring with an automorphism of a given group, the sum of the weights of the black vertices must be equal to a constant that depends on the colour of a given particular vertex. We study these colourings into four types of weighted cycles. We conclude this chapter with an application of the constant 2-labellings to the (r, a, b)-covering codes of the infinite grid with $|a - b| > 4$. We obtain the precise value of a and b for (r, a, b)-covering codes. Hence, we extend slightly the characterization obtained by Axenovich [Axe03].

We present in Appendix A the `Mathematica` code used to conjecture the recurrence relations satisfied by the 2-abelian complexity of the Thue–Morse word. The computation is based on matrix products that allows us to compute more efficiently the 2-abelian complexity than a more naive approach (like sliding a window of given length and then comparing the new factor to the ones already seen). For instance, the presented method only takes 4 minutes to compute the 65538 first terms on a standard computer (Intel® Core™ i3) while the naive approach already takes more than 4 minutes to compute the 513 first terms. Then, we give, through their graphics, the 1- and 2-abelian complexity sequences of pure morphic words over a 3-letter alphabet. Many of them exhibit a periodic behaviour or seem to satisfy a reflection symmetry. Finally, the proofs omitted in Chapter 2 for the period-doubling word are given in this appendix.

Appendix B contains the omitted proofs of Chapter 6 about constant 2-labellings in weighted cycles.

Conclusion

During my PhD, I have had the occasion to discover and to study some problems in combinatorics on words and covering problems in graphs. In addition to the work presented in this manuscript, I also studied the syntactic complexity associated with an eventually periodic set of integers represented in an integer base numeration system. With my co-advisor Rigo, a postdoctoral fellow Rampersad and a PhD student Lacroix, we obtained bounds on the number of elements in the syntactic monoid associated with these sets [LRRV12]. Generalizing Honkala's work [Hon86] and developing different techniques than the ones in [BCFR09], this study led to decision procedures on the periodicity of sets of integers given by finite automata.

The joint PhD offered me the opportunity to travel and to meet two communities of researchers at the different conferences and workshops. Also it allowed me to work with and to be part of several teams. I particularly enjoyed the scientific collaboration as attested by my nine co-authors of five distinct countries. The accomplished work gave rise to several publications (up to this day, two peer-reviewed publications in international revues [LRRV12, RSV13], one peer-reviewed publication in proceedings of an international conference [RV11] and three submitted articles [GPR$^+$, GV, PRRV]).

Part I

Combinatorics on words in an abelian context

Chapter 1

Words

This chapter contains all basic notions needed for a clear understanding of Part I. First, we recall some usual definitions and results about combinatorics on words. Secondly, we present three infinite words generated by morphisms: Fibonacci word, Thue–Morse word and period-doubling word. These words will be discussed again in the two following chapters. Next, we give definitions of automatic sequences and their generalization to regular sequences. Then, we consider the notion of ℓ-abelian complexity. Finally, we present a particular class of infinite binary words, called Sturmian words. We give characterizations of Sturmian words in terms of complexity, return words and rotation words.

Contents

1.1 Basic definitions about words

We briefly introduce the basic terminology on words. More details can be found in [BR10, Lot97]. An *alphabet* is a non-empty set A. We call *letters* its elements. A *word* over an alphabet A is a, finite or infinite, sequence of letters of A. Let ε denote the *empty word*, i.e., the empty sequence.

We denote by A^* the set of all finite words over A including the empty word and by $A^{\mathbb{N}}$ the set of (right) infinite words. The *length* of a word $w \in A^*$ is the number of its letters and is denoted by $|w|$. For a positive integer, we write A^n (respectively $A^{\leq n}$) the set of all words of length n (resp. of length at most n).

Remark 1.1. Our convention is that we index letters in an infinite word beginning with 0. Moreover, infinite words will always be denoted by a bold type letter, $\mathbf{w} \in A^{\mathbb{N}}$, while non-empty finite words are written with italic letters, $w \in A^*$. Since an infinite word is just a sequence over $\mathbb{N}$ taking values in a finite alphabet, we use the terms 'infinite word' and 'sequence' interchangeably.

If $u = u_0 \cdots u_{n-1}$ and $v = v_0 \cdots v_{m-1}$ are two finite words with $u_i, v_i \in A$, then the *concatenation* of u and v, denoted by $u \cdot v$ or simply uv, is given by

$$w = uv \text{ where } w_i = \begin{cases} u_i & \text{if } 0 \leq i < |u| \\ v_{i-|u|} & \text{if } |u| \leq i < |u| + |v|. \end{cases}$$

Note that the set A^* together with the concatenation form a monoid, where the empty word ε plays the role of the identity element.

We let u^n denote the concatenation of n copies of u and we set $u^0 = \varepsilon$. We denote by $u^\omega = uuu \cdots$ the infinite word obtained by concatenating copies of u. More precisely, the infinite word $\mathbf{w} = w_0 w_1 w_2 \cdots = u^\omega$ is defined by $w_i = u_{i \bmod n}$ for any $i \geq 0$.

We can also define the concatenation of a finite word $u = u_0 \cdots u_{n-1}$ with an infinite word $\mathbf{v} = v_0 v_1 v_2 \cdots$. The *concatenation* of u and $\mathbf{v}$, denoted by $u \cdot \mathbf{v}$ or simply $u\mathbf{v}$, is the infinite word $\mathbf{w}$ satisfying $w_i = u_i$ for any $i \in \{0, \ldots, |u| - 1\}$ and $w_i = v_{i-|u|}$ for any $i \geq |u|$.

Example 1.2. The concatenation of the words $fire$ and man gives the word $fireman$.

Let w be a finite word over A. A word u is called a *factor* (respectively *prefix*, *suffix*) of a word w if there exist two words x and y such that $w = xuy$ (resp. $w = uy$, $w = xu$). We denote respectively by $\operatorname{pref}_\ell(w)$ and $\operatorname{suff}_\ell(w)$ the prefix and suffix of length ℓ of w. The set of all factors (respectively prefixes, suffixes) of a word w is denoted by $\operatorname{Fac}(w)$ (resp. $\operatorname{Pref}(w)$, $\operatorname{Suff}(w)$).

These notions are similarly defined in the case of an infinite word $\mathbf{w} \in A^{\mathbb{N}}$. The sets of all the finite factors and of all finite prefixes of an infinite word $\mathbf{w}$ are given by

$$\operatorname{Fac}(\mathbf{w}) = \{u \in A^* \mid \exists x \in A^*, \mathbf{y} \in A^{\mathbb{N}}, \mathbf{w} = xu\mathbf{y}\},$$
$$\operatorname{Pref}(\mathbf{w}) = \{u \in A^* \mid \exists \mathbf{y} \in A^{\mathbb{N}}, \mathbf{w} = u\mathbf{y}\}.$$

Observe that we have $\operatorname{Pref}(\mathbf{w}) \subseteq \operatorname{Fac}(\mathbf{w})$. We define the sets $\operatorname{Pref}_n(\mathbf{w}) = \operatorname{Pref}(\mathbf{w}) \cap A^n$ and $\operatorname{Fac}_n(\mathbf{w}) = \operatorname{Fac}(\mathbf{w}) \cap A^n$, i.e., $\operatorname{Pref}_n(\mathbf{w})$ and $\operatorname{Fac}_n(\mathbf{w})$ are respectively the sets of prefixes and factors of length n of $\mathbf{w}$.

Let $i, j \in \mathbb{N}$ be such that $i \leq j$. The factor $w_i w_{i+1} \cdots w_j$ of $\mathbf{w} = w_0 w_1 \cdots$ is denoted by $\mathbf{w}[i, j]$. The notation $\mathbf{w}[i, i]$ is shortened to $\mathbf{w}_i$.

Example 1.3. Consider the word $\mathbf{w} = supercalifragilisticexpialidocious\cdots$. Then the word *super* is a prefix (and a factor) of $\mathbf{w}$. The word $fragilistic$ is a factor but not a prefix of $\mathbf{w}$. We have $\mathbf{w}[5,8] = cali$ and $\mathbf{w}_9 = f$.

Let $w \in A^*$. If a is a letter, $|w|_a$ denotes the number of occurrences of a in w. By analogy, for a finite word x, we let $|w|_x$ denote the number of occurrences of the factor x in u. For an ordered alphabet $A = \{a_1, \ldots, a_k\}$ with $a_1 < \cdots < a_k$, the *Parikh mapping* $\Psi : A^* \to \mathbb{N}^k$ is defined by

$$\Psi(w) = (|w|_{a_1}, \ldots, |w|_{a_k}).$$

The Parikh mapping is also known as the *commutative image*.

Example 1.4. Let $A = \{a, b\}$. The finite word $w = baaa$ has length $|w| = 4$ and it contains $|w|_a = 3$ occurrences of the letter a. We have $|w|_{aa} = 2$ and $|w|_{ab} = 0$. The last equality means that the word ab is not a factor of w. If we order the alphabet with the order $a < b$, then $\Psi(w) = (3, 1)$.

Let $C > 0$. An infinite word $\mathbf{w} \in A^\omega$ is C-*balanced*, if for all factors $u, v \in \mathrm{Fac}(\mathbf{w})$ of the same length, we have for all letters $a \in A$

$$||u|_a - |v|_a| \leq C.$$

A 1-balanced word is simply called *balanced*.

Example 1.5. Consider the infinite word $(\mathtt{01})^\omega = \mathtt{0101010101}\cdots$. Any factor of even length $2n$ has n zeroes and n ones, while any factor of odd length $2n + 1$ has n zeroes and $n + 1$ ones or vice versa. So the infinite word $(\mathtt{01})^\omega$ is 1-balanced.

If $w = w_0 \cdots w_{\ell-1}$ where w_i are letters, then we let $w^{\mathrm{R}} = w_{\ell-1} \cdots w_0$ denote the *reversal* of w. When a word is equal to its reversal, we say that the word is a *palindrome*.

Example 1.6. The reversal of the word $w = stressed$ is the word $w^{\mathrm{R}} = desserts$. The word $w = radar = w^{\mathrm{R}}$ is a palindrome since it is spelt the same backwards and forwards.

For any word $w = w_0 \cdots w_{\ell-1}$ of length ℓ over $\{\mathtt{0}, \mathtt{1}\}$, we write $\overline{w}$ for the *complement* of w, that is, the word obtained from w by changing $\mathtt{0}$'s into $\mathtt{1}$'s and $\mathtt{1}$'s into $\mathtt{0}$'s. Hence, the letters of $\overline{w}$ are given by $(\overline{w})_i = 1 - w_i$ for $0 \leq i \leq |w|$. Similarly, we let $\overline{\mathbf{w}}$ denote the complement of $\mathbf{w} \in \{\mathtt{0}, \mathtt{1}\}^{\mathbb{N}}$, in the case of infinite words.

Example 1.7. The complement of $w = \mathtt{011010}$ is the word $\overline{w} = \mathtt{100101}$.

If a word w starts with the letter a, then $a^{-1}w$ denotes the word obtained from w by deleting its first letter. Similarly, if a word w ends with the letter a, then wa^{-1} denotes the word obtained from w by deleting its last letter.

Example 1.8. If w is the word $tear$, then $t^{-1}w = ear$ and $wr^{-1} = tea$.

An infinite word $\mathbf{w}$ is *periodic (of period m)*, if it can be factored as $\mathbf{w} = u^\omega = uuu\cdots$ with $u \in A^*$ and $|u| = m$. The smallest m for which such a factorization exists is called the *period* of $\mathbf{w}$. An infinite word $\mathbf{w}$ is called *ultimately periodic* if it is the concatenation of a finite word with a periodic infinite word. An infinite word $\mathbf{w}$ is *aperiodic* if it is not ultimately periodic.

Example 1.9 (Representation of reals)**.** Real numbers from the interval $[0,1)$ are traditionally represented as infinite sequences of digits, that is as infinite words. Let k be a positive integer and $\alpha \in [0,1)$. The *k-expansion* of α is the infinite word $a_1 a_2 \cdots$ over the alphabet $\{\mathtt{0}, \dots, k-1\}$ such that the letters a_i satisfy

$$\alpha = \sum_{i=0}^{\infty} a_i k^{-i}.$$

It is well known that a number is rational if and only if its k-expansion is eventually periodic [HW08]. For instance, we can check that the 10-expansion of $11/28$, i.e., the decimal digits of $11/28$, is the infinite word $\mathtt{39}(\mathtt{28514})^\omega$ as

$$\frac{11}{28} = 0.392857142857142857142857142857142857142857142 \cdots.$$

An infinite word $\mathbf{w}$ is said to be *recurrent* if every factor u of $\mathbf{w}$ appears infinitely often in $\mathbf{w}$. An infinite word $\mathbf{w}$ is said to be *uniformly recurrent*, if it is recurrent and the distance between any two consecutive occurrences of any of its factors u is bounded by a constant depending only on u. In other words, $\mathbf{w}$ is uniformly recurrent if it is recurrent and for all $u \in \mathrm{Fac}(\mathbf{w})$, there exists a constant $C_u > 0$ such that for all $i \geq 0$, we have

$$\mathbf{w}[i, i+|u|-1] = u \Rightarrow (\exists j \in \{i+1, \dots, i+C_u\} : \mathbf{w}[j, j+|u|-1] = u).$$

Many notions occurring in combinatorics on words have been recently and fruitfully extended to an abelian context. Two words u and v are said to be *abelian equivalent* if u is an anagram of v. For instance, the words *silent* and *listen* are abelian equivalent.

Definition 1.10. Let u, v be two finite words of the same length. We say that u and v are *abelian equivalent* and we write $u \sim_{ab} v$ if $\Psi(u) = \Psi(v)$. For example, $\mathtt{00011} \sim_{ab} \mathtt{10010}$. Note that we see in Section 1.4 that the notion of abelian equivalence can be generalized to ℓ-abelian equivalence.

Usually, the corresponding concepts are defined up to abelian equivalence. The notion of avoidance in an abelian context was introduced by Erdős in the 60's [Erd61]. He raised the question whether there exist infinite words over a given alphabet that do not contain *abelian squares*, i.e., two consecutive permutations of the same factor. The word *intestines* is an example of abelian square. Since then many authors have considered questions such as avoiding abelian repetitions. For instance, abelian squares are avoidable on a 4-letter alphabet [Ker92]. Abelian cubes are avoidable on a 3-letter alphabet [Dek79]. Abelian 4^{th} power are avoidable on a 2-letter alphabet [Dek79].

An infinite word $\mathbf{w}$ is *abelian periodic (of period m)*, if it can be factored as $\mathbf{w} = u_1 u_2 u_3 \cdots$ where, for all i, j, the finite words u_i and u_j have the same length m and are abelian equivalent. The smallest m for which such a factorization exists is called the *abelian period* of $\mathbf{w}$. An infinite word $\mathbf{w}$ is called *ultimately abelian periodic* if it is the concatenation of a finite word with an abelian periodic infinite word.

Example 1.11 (Thue–Morse word)**.** Consider the infinite word $\mathbf{t} = t_0 t_1 t_2 \cdots$ over the alphabet $\{\mathtt{0}, \mathtt{1}\}$ called the *Thue–Morse word* [AS99] and defined by the recurrence relations

$$t_0 = \mathtt{0},\ t_{2i} = t_i \text{ and } t_{2i+1} = \mathtt{1} - t_i \text{ for all } i \in \mathbb{N}.$$

The prefix of length 32 of $\mathbf{t}$ is then given by

$$\mathrm{pref}_{32}(\mathbf{t}) = \mathtt{01101001100101101001011001101001}.$$

The Thue–Morse word $\mathbf{t}$ appears as A010060 in [OF]. We have $t_{2i+1} = \overline{t_{2i}}$ for all $i \geq 0$. So $\mathbf{t}$ is the concatenation of $\mathtt{01}$ and $\mathtt{10}$. Since these words $\mathtt{01}$ and $\mathtt{10}$ are abelian equivalent, $\mathbf{t}$ is abelian periodic of period 2.

Let $\mathbf{w}$ be an infinite word. If, for each factor u of $\mathbf{w}$, there exist infinitely many i such that $\mathbf{w}[i, i+|u|-1] \sim_{ab} u$, then $\mathbf{w}$ is said to be *abelian recurrent*. If $\mathbf{w}$ is abelian recurrent and if, for each factor u of $\mathbf{w}$, the distance between any two consecutive occurrences of factors abelian equivalent to u is bounded by a constant depending only on u, then $\mathbf{w}$ is said to be *abelian uniformly recurrent*.

Remark 1.12. Note that uniform recurrence implies obviously abelian uniform recurrence. We show in Chapter 3 that the converse does not hold.

1.2 Infinite words generated by morphisms

In this thesis, we consider three notable infinite words : the Fibonacci word, the Thue–Morse word and the period-doubling word. We can define these words using morphisms.

Let A and B be two alphabets. A *morphism* is a map $\varphi : A^* \to B^*$ that satisfies $\varphi(uv) = \varphi(u)\varphi(v)$ for all $u, v \in A^*$. Typically we use the Greek letters $\varphi, \eta, \theta, \mu, \nu, \sigma$ to denote morphisms.

Clearly for any morphism φ, we have $\varphi(\varepsilon) = \varepsilon$. Moreover, a morphism is completely determined by the images of the letters of the alphabet A. Hence, when we define a morphism, we always give it by specifying these images. If all these images have the same length k, then we say that the morphism is *k-uniform*. A 1-uniform morphism is called a *coding* or a *letter-to-letter* morphism.

Example 1.13. Let $A = \{a, e, r, t\}$, $B = \{e, i, \ell, m, s\}$, $C = \{f, n, u, y\}$ and define the morphisms

$$\varphi : A^* \to B^*, \begin{cases} a \mapsto \varepsilon \\ e \mapsto mi \\ r \mapsto \ell e \\ t \mapsto s \end{cases} \quad \text{and} \quad \tau : B^* \to C^*, \begin{cases} e \mapsto y \\ i \mapsto n \\ \ell \mapsto n \\ m \mapsto u \\ s \mapsto f \end{cases}.$$

Then $\varphi(tear) = smi\ell e$ and $\tau(smi\ell e) = funny$. Moreover τ is an example of coding.

Consider now that $A = B$. Hence, consider a morphism $\varphi : A^* \to A^*$. Then we can iterate the application of φ. We define $\varphi^0(a) = a$ and $\varphi^{i+1}(a) = \varphi(\varphi^i(a))$ for all $a \in A$. The morphism φ is called *primitive* if there exists a positive integer k such that $|\varphi^k(a)|_b \geq 1$ for all letters $a, b \in A$. We can also apply morphisms to infinite words. If $\mathbf{w} = w_0 w_1 w_2 \cdots$, then we define

$$\varphi(\mathbf{w}) = \varphi(w_0)\varphi(w_1)\varphi(w_2)\cdots.$$

An infinite word $\mathbf{w}$ such that $\varphi(\mathbf{w}) = \mathbf{w}$ is called a *fixed point* of φ.

We equip the set $A^{\mathbb{N}}$ of a *distance* turning it into a metric space[1]. Let $\mathbf{w}, \mathbf{x}$ be two distinct infinite words in $A^{\mathbb{N}}$. We denote by $\mathbf{w} \wedge \mathbf{x}$ the longest common prefix of $\mathbf{w}$ and $\mathbf{x}$. We define a distance $d : A^{\mathbb{N}} \times A^{\mathbb{N}} \to [0, \infty)$ by

$$d(\mathbf{w}, \mathbf{x}) = 2^{-|\mathbf{w} \wedge \mathbf{x}|}$$

and we set $d(\mathbf{w}, \mathbf{w}) = 0$. Hence, the longer prefix two words share, the closer they are. We can now introduce the notion of a sequence of finite words converging to an infinite word. Let $(u_n)_{n \geq 0}$ be a sequence of finite words over A. If \$ is a symbol that does not belong to A, then any word $u \in A^*$ is in bijection with the word $u\$^\omega \in (A \cup \{\$\})^{\mathbb{N}}$. So we say that the sequence $(u_n)_{n \geq 0}$ of finite words is *converging* to the infinite word $\mathbf{w}$ if the sequence $(u_n\$^\omega)_{n \geq 0}$ of infinite words is converging to $\mathbf{w}$, i.e., if

$$\lim_{n \to \infty} d(\mathbf{w}, u_n\$^\omega) = 0.$$

Let a be a letter of A. If there exists a finite word $u \in A^*$ such that $\varphi(a) = au$ and $\lim_{n \to +\infty} |\varphi^n(a)| = +\infty$, then φ is *prolongable* on the letter a. Note that if φ is prolongable on a letter, then we say that φ is a *substitution*. In that case, the sequence of words $a, \varphi(a), \varphi^2(a), \ldots$ converges to the infinite word

$$\varphi^\omega(a) = au\varphi(u)\varphi^2(u) \cdots ,$$

which is a fixed point of φ as $\varphi(\varphi^\omega(a)) = \varphi^\omega(a)$. If $\mathbf{w} = \varphi^\omega(a)$, we say that $\mathbf{w}$ is a *(pure) morphic word* or a *substitutive word*. Let C be an alphabet. If $\mathbf{w} = \tau(\varphi^\omega(a))$ where τ is a coding from A^* to C^*, then $\mathbf{w}$ is a *morphic word*. If moreover the morphism φ is primitive, then we say that $\mathbf{w}$ is a *primitive morphic word*.

Let us now introduce three notable examples of morphic words. The first two examples are discussed again in Chapter 3.

Example 1.14 (Fibonacci word)**.** Let $A = \{\mathtt{0}, \mathtt{1}\}$. The *Fibonacci morphism* $\varphi : A^* \to A^*$ is defined as follows

$$\varphi : \begin{cases} \mathtt{0} \mapsto \mathtt{01} \\ \mathtt{1} \mapsto \mathtt{0}. \end{cases}$$

The first few iterations of φ are $\varphi(\mathtt{0}) = \mathtt{01}$, $\varphi^2(\mathtt{0}) = \mathtt{010}$, $\varphi^3(\mathtt{0}) = \mathtt{01001}$, $\varphi^4(\mathtt{0}) = \mathtt{01001010}$. It is easy to check that $|\varphi^n(\mathtt{0})|$ is the n-th Fibonacci number. Hence $\varphi^\omega(\mathtt{0})$ is well-defined and the *Fibonacci word* is the fixed point of the morphism φ starting with $\mathtt{0}$:

$$\mathbf{f} = \varphi^\omega(\mathtt{0}) = \mathtt{0100101001001010010100100101001001} \cdots .$$

This word appears as A003849 in [OF].

Example 1.15. The Thue–Morse word $\mathbf{t}$ introduced in Example 1.11 is also a fixed point of a morphism. Consider the morphism σ over the free monoid $\{\mathtt{0}, \mathtt{1}\}^*$ defined by

$$\sigma : \begin{cases} \mathtt{0} \mapsto \mathtt{01} \\ \mathtt{1} \mapsto \mathtt{10}. \end{cases}$$

It is easy to check that the *Thue–Morse word*

$$\mathbf{t} = \sigma^\omega(\mathtt{0}) = \mathtt{0110100110010110100101100110100l} \cdots$$

is the fixed point of σ starting with $\mathtt{0}$.

[1] In fact, the set $A^{\mathbb{N}}$ equipped with the distance will even be an ultrametric space.

Example 1.16 (Period-doubling word)**.** The *period-doubling word*, known as the sequence A096268 in [OF],

$$\mathbf{p} = \psi^\omega(\mathtt{0}) = \mathtt{010001010100010001000}\cdots$$

is a fixed point of the morphism ψ over $\{\mathtt{0},\mathtt{1}\}^*$ defined by

$$\psi : \begin{cases} \mathtt{0} \mapsto \mathtt{01} \\ \mathtt{1} \mapsto \mathtt{00}. \end{cases}$$

We consider again the last two examples in Chapter 2. Observe that these last two sequences are obtained by iterating a 2-uniform morphism. These sequences are in fact two 2-automatic sequences as shown in the next section.

1.3 Automatic and regular sequences

Let $k \geq 2$ be an integer. A *k-automatic sequence* is an infinite word $\mathbf{w} = \tau(\varphi^\omega(a))$, where φ is a k-uniform morphism, τ is a coding and a is a letter. Since the fundamental work of Cobham [Cob72], the automatic sequences have been extensively studied. See for instance [Eil74, AS03a].

Example 1.17. The Thue–Morse word $\mathbf{t}$ and the period-doubling word $\mathbf{p}$ are both generated by 2-uniform morphisms (see Example 1.15 and 1.16). Hence, they are both 2-automatic.

We see that the automatic sequences are related to base-k representations, also known as base-k expansions. Let n be a positive integer. The *base-k representation* of n is the word $a_\ell a_{\ell-1} \cdots a_0$ over the alphabet $\{\mathtt{0},\ldots,k-\mathtt{1}\}$ where the letters a_i satisfy

$$n = \sum_{i=0}^{\ell} a_i k^i$$

and $a_\ell \neq 0$. We denote the base-k representation of n by $\mathrm{rep}_k(n)$ and we say that n is the *value* of the word $\mathrm{rep}_k(n) = a_\ell a_{\ell-1} \cdots a_0$. Moreover, this definition implies that numbers are written with the *most significant digit* on the left. For instance, the base-3 representation of $n = 11$ is the word $\mathrm{rep}_3(11) = \mathtt{102}$ over the alphabet $\{\mathtt{0},\mathtt{1},\mathtt{2}\}$ since $11 = 1 \cdot 9 + 0 \cdot 3 + 2 \cdot 1$.

We give some basic properties of automatic sequences. These results can be found in [AS03a].

Theorem 1.18. [AS03a]

If a sequence $\mathbf{w}$ differs only in finitely many terms from a k-automatic sequence, then it is also k-automatic.

Theorem 1.19. [AS03a]

If $\mathbf{w}$ is an ultimately periodic sequence, then it is k-automatic for all $k \geq 2$.

Theorem 1.20. [AS03a]

Let $\mathbf{w}$ be a k-automatic sequence and τ be a coding. Then the sequence $\tau(\mathbf{w})$ is also k-automatic.

Automatic sequences can be characterized by their k-kernel. The *k-kernel* of a sequence $\mathbf{s} = s(n)_{n\geq 0}$ is the set

$$\mathcal{K}_k(\mathbf{s}) = \{s(k^e n + r)_{n\geq 0} : e \geq 0 \text{ and } 0 \leq r < k^e\}.$$

Theorem 1.21. [Eil74, AS03a]

A sequence $\mathbf{w}$ is k-automatic if and only if its k-kernel is finite.

Example 1.22. For instance, the 2-kernel $\mathcal{K}_2(\mathbf{t})$ of the Thue–Morse word $\mathbf{t}$ contains exactly two elements, namely $\mathbf{t}$ and $\overline{\mathbf{t}} = \sigma^\omega(\mathtt{1})$, since $\sigma(\mathtt{0}) = \mathtt{0}\overline{\mathtt{0}}$ and $\sigma(\mathtt{1}) = \mathtt{1}\overline{\mathtt{1}}$. In other words, for $\mathbf{t} = t_0 t_1 t_2 \cdots$, we have $t_{2i} = t_i$ and $t_{2i+1} = \overline{t_i}$.

Remark 1.23. Another characterization of k-automatic sequences involves automaton with output. An automaton with output is just a finite set of elements represented by circles, that are linked with labelled arcs, and inside each of these elements, there is an output.

More formally, a *deterministic finite automaton with output* (DFAO) is a 6-tuple $\mathcal{M} = (Q, q_0, A, \delta, B, \mu)$ where Q is a non-empty finite set of elements called *states*, $q_0 \in Q$ is the *initial state*, A is an alphabet, $\delta : Q \times A \to Q$ is a *transition function*, B is an *output alphabet* and $\mu : Q \to B$ is an *output function*. The transition function δ naturally extends to a function on $Q \times A^*$ as follows

$$\delta(q, \varepsilon) = q \text{ and } \delta(q, aw) = \delta(\delta(q, a), w)$$

for any state $q \in Q$, letter $a \in A$ and word $w \in A^*$.

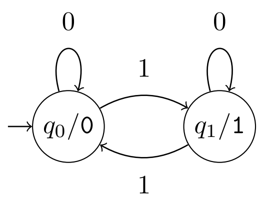


Figure 1.1: A deterministic finite automaton with output (DFAO).

For example, Figure 1.1 depicts a DFAO with a set of states $Q = \{q_0, q_1\}$, q_0 as its initial state (represented with a small incoming arc), an alphabet $A = \{0, 1\}$, a function transition δ defined by

$$\delta(q_0, 0) = q_0,\ \delta(q_0, 1) = q_1,\ \delta(q_1, 0) = q_1 \text{ and } \delta(q_1, 1) = q_0,$$

an output alphabet $B = \{\mathtt{0}, \mathtt{1}\}$ and an output function μ defined by $\mu(q_0) = \mathtt{0}$ and $\mu(q_1) = \mathtt{1}$.

Theorem 1.24. [Cob72]

Let $\mathbf{w} = w_0 w_1 w_2 \cdots$ be an infinite word over an alphabet B. It is of the form $\mathbf{w} = \tau(\varphi^\omega(a))$, where $\varphi : A^* \to A^*$ is a k-uniform morphism, $\tau : A^* \to B^*$ is a coding and a is a letter of A if and only if there exists a DFAO $(Q, q_a, \{0, \ldots, k-1\}, \delta, B, \mu)$ such that $Q = \{q_i : i \in A\}$, $\delta(q_a, 0) = q_a$ and

$$w_j = \mu(\delta(q_a, \mathrm{rep}_k(j))) \quad \forall j \geq 0.$$

Example 1.25. We already know that the Thue–Morse word $\mathbf{t}$ is 2-automatic since it is generated by a 2-uniform morphism. In addition, we can check that the DFAO in Figure 1.1 is generating $\mathbf{t}$ when we feed him with the base-2 representation of the indices (Table 1.1). To do this, observe first that the definition given in Example 1.11 implies

$$t_i = \begin{cases} \mathtt{1} & \text{if } |\mathrm{rep}_2(i)|_1 \text{ is even} \\ \mathtt{0} & \text{otherwise.} \end{cases}$$

Then for instance, consider t_6. The base-2 representation of 6, $\mathrm{rep}_2(6) = 110$, is processed from most significant to least significant digit: $q_0 \xrightarrow{1} q_1 \xrightarrow{1} q_0 \xrightarrow{0} q_0$. Hence the automaton produces the output associated with the state q_0 which is $\mathtt{0}$. So we find $t_6 = \mathtt{0}$.

i	0	1	2	3	4	5	6	7	8	...
t_i	0	1	1	0	1	0	0	1	1	...
$\mathrm{rep}_2(i)$	ε	1	10	11	100	101	110	111	1000	...

Table 1.1: The base-2 representations of the first positive integers and the corresponding letters of the Thue–Morse word $\mathbf{t}$.

A natural generalization of automatic sequences to sequences over an infinite alphabet is given by the notion of k-regular sequences, introduced by Allouche and Shallit [AS92]. The k-regularity of a sequence provides us with structural information about how the different terms are related to each other. We restrict ourselves to sequences taking integer values only.

Definition 1.26. Let $k \geq 2$ be an integer. A sequence $\mathbf{s} = s(n)_{n\geq 0} \in \mathbb{Z}^{\mathbb{N}}$ is *k-regular* if the ideal $\langle \mathcal{K}_k(\mathbf{s}) \rangle$ generated by $\mathcal{K}_k(\mathbf{s})$ is a finitely-generated $\mathbb{Z}$-module, i.e., there exist a finite number of sequences $t_1(n)_{n\geq 0}, \ldots, t_\ell(n)_{n\geq 0}$ such that every sequence in the k-kernel $\mathcal{K}_k(\mathbf{s})$ is a $\mathbb{Z}$-linear combination of the t_i's. Otherwise stated, for all $e \geq 0$ and for all $r \in \{0, \ldots, k^e - 1\}$, there exist integers $c_1, \ldots, c_\ell$ such that

$$\forall n \geq 0, \quad s(k^e n + r) = \sum_{i=1}^{\ell} c_i \, t_i(n).$$

Clearly, from Theorem 1.21, any k-automatic sequence is necessarily k-regular. There are many natural examples of k-regular sequences [AS92, AS03b].

Example 1.27 (Sums of digits)**.** Consider the sequence $s(n)_{n\geq 0}$ where $s(n)$ is the sum of the digits of the base-2 representation of n:

$$s(n)_{n\geq 0} = (0,1,1,2,1,2,2,3,1,2,2,3,2,3,3,4,1,2,2,3,2,3,3,4,2,3,3,4,3,\ldots)$$

Then

$$s(2^e n + r) = s(n) + s(r)$$

for all $e \geq 0$ and for all $r \in \{0,\ldots,k^e - 1\}$, since $\mathrm{rep}_2(2^e n) = \mathrm{rep}_2(n)0^e$ and $|\mathrm{rep}_2(r)| \leq e$. Therefore every sequence in the 2-kernel is a $\mathbb{Z}$-linear combination of the sequence $s(n)_{n\geq 0}$ and the constant sequence 1. So the 2-kernel is finitely generated and $s(n)_{n\geq 0}$ is a 2-regular sequence.

There is a convenient matrix representation for k-regular sequences which leads to an efficient algorithm for computing the values of such a sequence (and many related quantities). See also [BR11, Chapter 5] for connections with rational series.

Theorem 1.28. [AS92]

A sequence $s(n)_{n\geq 0}$ is k-regular if and only if there exist ℓ sequences $s = s_1, s_2, \ldots, s_\ell$ and k matrices $B_0, B_1, \ldots, B_{k-1}$ of size $\ell \times \ell$ for some integer ℓ such that if

$$V(n) = \begin{pmatrix} s_1(n) \\ \vdots \\ s_\ell(n) \end{pmatrix},$$

one has $V(kn + a) = B_a V(n)$ for $0 \leq a < k$.

Example 1.29 (Example 1.27 continued)**.** For the sequence $s(n)_{n\geq 0}$ of the sum of digits of the base-2 representation, we set $\ell = 2$, $k = 2$,

$$V(n) = \begin{pmatrix} s(n) \\ 1(n) \end{pmatrix}, B_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } B_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

where $1(n)_{n\geq 0}$ denotes the constant sequence 1. Then, we have $V(2n) = B_0 V(n)$ and $V(2n+1) = B_1 V(n)$. This shows that $s(n)_{n\geq 0}$ is a 2-regular sequence.

In Chapter 2, we use the following classical results stated as Theorem 2.3, Corollary 2.4 and Theorem 2.5 in [AS92].

Lemma 1.30. [AS92]

Let $k \geq 2$ be an integer. A sequence taking finitely many values is k-regular if and only if it is k-automatic.

Lemma 1.31. [AS92]

Let $k, m \geq 2$ be integers. If a sequence $s(n)_{n\geq 0}$ is k-regular, then $(s(n) \bmod m)_{n\geq 0}$ is k-automatic.

Lemma 1.32. [AS92]

Let $k \geq 2$ be an integer. If the sequences $s_1(n)_{n\geq 0}$ and $s_2(n)_{n\geq 0}$ are k-regular, then so are the sequences

$$(s_1(n) + s_2(n))_{n\geq 0},\ (s_1(n) \cdot s_2(n))_{n\geq 0} \text{ and } (c \cdot s_1(n))_{n\geq 0}$$

for any constant c.

Moreover, as a direct consequence of two results stated in [AS92], namely Theorem 2.6 and its following remark, we have the next lemma.

Lemma 1.33.

Let $k \geq 2$ be an integer. Let $s(n)_{n\geq 0}$ be a sequence. The sequence $s(n)_{n\geq 0}$ is k-regular if and only if $s(n+1)_{n\geq 0}$ is k-regular.

We often make use of the following composition lemma for a function F defined piecewise on several k-automatic sets. This lemma is a direct consequence of Lemma 1.32.

Lemma 1.34.

Let $k \geq 2$. Let $P_1, \ldots, P_\ell : \mathbb{N} \to \{0, 1\}$ be unary predicates that are k-automatic. Let $f_1, \ldots, f_\ell$ be k-regular functions. The function $F : \mathbb{N} \to \mathbb{N}$ defined by

$$F(n) = \sum_{i=1}^{\ell} f_i(n)\, P_i(n)$$

is k-regular.

Note that if, for each n, there is exactly one i such that $P_i(n) = 1$, then we can write

$$F(n) = \begin{cases} f_1(n) & \text{if } P_1(n) = 1 \\ f_2(n) & \text{if } P_2(n) = 1 \\ \quad\vdots & \qquad\vdots \\ f_\ell(n) & \text{if } P_\ell(n) = 1. \end{cases}$$

This is the setting in which we apply Lemma 1.34.

1.4 Factor complexity and ℓ-abelian complexity

Given an infinite word, it is natural to wish to determine how "complex" the word is. They are many measures of complexity. We are interested in the ones counting factors up to some equivalence.

A classical measure of the complexity of an infinite word $\mathbf{w}$ is its *factor complexity* $\mathcal{P}^{(\infty)}_{\mathbf{w}} : \mathbb{N} \to \mathbb{N}$ which maps n to the number of distinct factors of length n occurring in $\mathbf{w}$.

Example 1.35. The factor complexity $\mathcal{P}^{(\infty)}_{\mathbf{t}}$ of the Thue–Morse word $\mathbf{t}$ is well-known [Brl89, dLV88]. The first values are depicted in Figure 1.2. We have $\mathcal{P}^{(\infty)}_{\mathbf{t}}(0) = 1$, $\mathcal{P}^{(\infty)}_{\mathbf{t}}(1) = 2$, $\mathcal{P}^{(\infty)}_{\mathbf{t}}(2) = 4$ and for $n \geq 2$,

$$\mathcal{P}^{(\infty)}_{\mathbf{t}}(n) = \begin{cases} 4n - 2 \cdot 2^m - 4 & \text{if } 2 \cdot 2^m < n \leq 3 \cdot 2^m \\ 2n + 4 \cdot 2^m - 2 & \text{if } 3 \cdot 2^m < n \leq 4 \cdot 2^m. \end{cases}$$

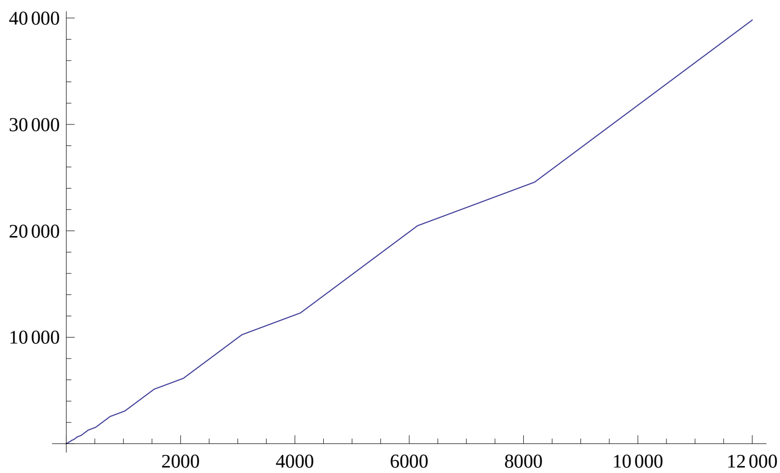


Figure 1.2: The values of the factor complexity $\mathcal{P}^{(\infty)}_{\mathbf{t}}(n)$ of the Thue–Morse word for n in $\{0, \ldots, 12000\}$.

The well-known theorem of Morse–Hedlund gives a characterization of the ultimately periodic words in terms of factor complexity. See for instance [Lot97, Theorem 2.1.5] or [AS03a, Theorem 10.2.6] for a proof.

Theorem 1.36. Morse–Hedlund theorem

An infinite word $\mathbf{w}$ is ultimately periodic if and only if there exists a positive integer n such that the factor complexity of $\mathbf{w}$ satisfies $\mathcal{P}^{(\infty)}_{\mathbf{w}}(n) \leq n$.

The next result was obtained independently by Charlier et al. [CRS12] and by Carpi and D'Alonzo [CD10]. The latter authors actually proved that the result holds for a larger class of sequences, called k-synchronized sequences. The special case of sequences generated by uniform morphisms (without applying a coding after) was also proved by Mossé [Mos96].

Proposition 1.37. [CD10, CRS12]

A k-automatic sequence $\mathbf{w}$ has a k-regular factor complexity function and the sequence $(\mathcal{P}_{\mathbf{w}}^{(\infty)}(n+1) - \mathcal{P}_{\mathbf{w}}^{(\infty)}(n))_{n\geq 0}$ is k-automatic.

Example 1.38. Consider again the Thue–Morse word. We have

$$\mathcal{P}_{\mathbf{t}}^{(\infty)}(2n+1) = 2\mathcal{P}_{\mathbf{t}}^{(\infty)}(n+1) \text{ and } \mathcal{P}_{\mathbf{t}}^{(\infty)}(2n) = \mathcal{P}_{\mathbf{t}}^{(\infty)}(n+1) + \mathcal{P}_{\mathbf{t}}^{(\infty)}(n)$$

for all $n \geq 2$. It is not obvious from these relations that the factor complexity is 2-regular, but we also have the following relations

$$\begin{aligned}
\mathcal{P}_{\mathbf{t}}^{(\infty)}(8n) &= -2\mathcal{P}_{\mathbf{t}}^{(\infty)}(2n) + 3\mathcal{P}_{\mathbf{t}}^{(\infty)}(4n)\\
\mathcal{P}_{\mathbf{t}}^{(\infty)}(4n+1) &= -\mathcal{P}_{\mathbf{t}}^{(\infty)}(2n) + \mathcal{P}_{\mathbf{t}}^{(\infty)}(2n+1) + \mathcal{P}_{\mathbf{t}}^{(\infty)}(4n)\\
\mathcal{P}_{\mathbf{t}}^{(\infty)}(4n+3) &= \mathcal{P}_{\mathbf{t}}^{(\infty)}(2n) - \mathcal{P}_{\mathbf{t}}^{(\infty)}(2n+1) - \mathcal{P}_{\mathbf{t}}^{(\infty)}(4n) + 2\mathcal{P}_{\mathbf{t}}^{(\infty)}(4n+2)\\
\mathcal{P}_{\mathbf{t}}^{(\infty)}(8n+2) &= -2\mathcal{P}_{\mathbf{t}}^{(\infty)}(2n) + 2\mathcal{P}_{\mathbf{t}}^{(\infty)}(4n) + \mathcal{P}_{\mathbf{t}}^{(\infty)}(4n+2)\\
\mathcal{P}_{\mathbf{t}}^{(\infty)}(8n+4) &= \mathcal{P}_{\mathbf{t}}^{(\infty)}(2n) - \mathcal{P}_{\mathbf{t}}^{(\infty)}(2n+1) - \mathcal{P}_{\mathbf{t}}^{(\infty)}(4n) + 3\mathcal{P}_{\mathbf{t}}^{(\infty)}(4n+2)\\
\mathcal{P}_{\mathbf{t}}^{(\infty)}(8n+6) &= -2\mathcal{P}_{\mathbf{t}}^{(\infty)}(n) + 7\mathcal{P}_{\mathbf{t}}^{(\infty)}(2n) - 2\mathcal{P}_{\mathbf{t}}^{(\infty)}(2n+1) - 5\mathcal{P}_{\mathbf{t}}^{(\infty)}(4n) + 5\mathcal{P}_{\mathbf{t}}^{(\infty)}(4n+2)
\end{aligned}$$

which can be checked using the closed formula given in Example 1.35.

In an abelian context, the analogue to factor complexity is the abelian complexity. The *abelian complexity* of an infinite word $\mathbf{w}$ is a function $\mathcal{P}_{\mathbf{w}}^{(1)} : \mathbb{N} \to \mathbb{N}$ which maps n to the number of distinct factors of length n occurring in $\mathbf{w}$, counted up to abelian equivalence. For more details on abelian complexity, see the reference [RSZ11] which contains many relevant bibliographic pointers.

Example 1.39. The abelian complexity of the Thue–Morse word satisfies the following equalities

$$\mathcal{P}_{\mathbf{t}}^{(1)}(2n+1) = 2 \text{ and } \mathcal{P}_{\mathbf{t}}^{(1)}(2n) = 3.$$

Indeed, for a factor u of odd length, we have $|u|_{\mathtt{0}} = |u|_{\mathtt{1}}+1$ or $|u|_{\mathtt{0}} = |u|_{\mathtt{1}}-1$. Both possibilities occur: if $u \in \text{Fac}(\mathbf{t})$, then we also have $\overline{u} \in \text{Fac}(\mathbf{t})$ since $\sigma^{n+1}(\mathtt{0}) = \sigma^n(\mathtt{0})\sigma^n(\overline{\mathtt{0}})$. Now for a factor u of even length, either it occurs at an even index and $|u|_{\mathtt{0}} = |u|_{\mathtt{1}}$ by definition of σ, or it occurs at an odd index. In the last case (Figure 1.3), we can write u as the concatenation of $u_0 v_1 \cdots v_{n-1} u_{2n-1}$ where $u_0, u_{2n-1} \in \{\mathtt{0}, \mathtt{1}\}$ and $v_1, \ldots, v_{n-1} \in \{\mathtt{0}, \mathtt{1}\}^2$. There are three possible abelian equivalence classes depending on the first letter u_0 and the last letter u_{2n-1} of u :

$$\begin{cases} |u|_{\mathtt{0}} = |u|_{\mathtt{1}} - 2 & \text{if } u_0 = u_{2n-1} = \mathtt{1}\\ |u|_{\mathtt{0}} = |u|_{\mathtt{1}} & \text{if } u_0 \neq u_{2n-1}\\ |u|_{\mathtt{0}} = |u|_{\mathtt{1}} + 2 & \text{if } u_0 = u_{2n-1} = \mathtt{0}. \end{cases}$$

Again all cases actually occur.

Bounded factor complexity can be interpreted in terms of ultimate periodicity (see the Morse–Hedlund theorem). Similarly, bounded abelian complexity can be interpreted as follows.

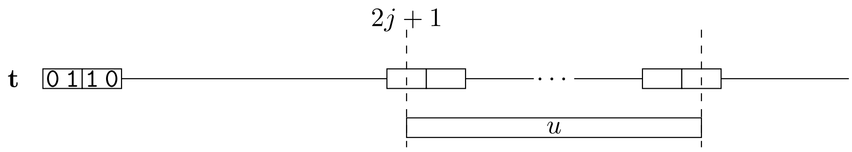


Figure 1.3: A factor u of even length occurring at an odd index $2j+1$ starts and ends at the middle of a factor `01` or `10`.

Lemma 1.40. [RSZ11]

An infinite word has bounded abelian complexity if and only if it is C-balanced for some $C > 0$.

Let $\ell \geq 1$ be an integer. Based on [Kar80] the notions of abelian equivalence and thus abelian complexity were recently extended to ℓ-abelian equivalence and ℓ-abelian complexity [KSZ13].

Definition 1.41. Let u, x be two finite words. Recall that $|u|_x$ denote the number of occurrences of the factor x in u. Two finite words u and v are *ℓ-abelian equivalent* if $|u|_x = |v|_x$ for all words x of length $|x| \leq \ell$. In that case, we write $u \sim_{ab,\ell} v$.

Example 1.42. The words `11010011` and `11101001` are 2-abelian equivalent (see Table 1.2) but they are not 3-abelian equivalent. Indeed, the factor `011` occurs in the first word but not in the second one.

u	$\lvert u\rvert_{0}$	$\lvert u\rvert_{1}$	$\lvert u\rvert_{00}$	$\lvert u\rvert_{01}$	$\lvert u\rvert_{10}$	$\lvert u\rvert_{11}$
11010011	3	5	1	2	2	2
11101001	3	5	1	2	2	2

Table 1.2: The number of occurrences of factors of length at most 2 into the words `11010011` and `11101001`.

The following lemma gives an equivalent way of defining the ℓ-abelian equivalence.

Lemma 1.43. [KSZ13]

Let u, v be words of length at least $\ell - 1$ and let $|u|_x = |v|_x$ for every word x of length ℓ. The following conditions are equivalent:

- $|u|_x = |v|_x$ for all $x \in A^{\leq \ell - 1}$,
- $\mathrm{pref}_{\ell-1}(u) = \mathrm{pref}_{\ell-1}(v)$,
- $\mathrm{suff}_{\ell-1}(u) = \mathrm{suff}_{\ell-1}(v)$.

The question of avoidance was extended to the context of ℓ-abelian equivalence. Table 1.3 depicts the answer to the following question for $n, m \in \{2, 3, 4\}$. *Given $n, m \geq 2$, what is the smallest ℓ such that ℓ-abelian n^{th} powers are avoidable on an m-letter alphabet?* For instance, the size of the smallest alphabet where 2-abelian squares are avoidable is four [HKS12], like in the case of the abelian squares [Ker92]. This size is two for 2-abelian cubes [Rao], like in the case of "usual" cubes [Thu06, Thu12].

$n \diagdown m$	2	3	4
2	-	3 [Rao]	1 [Ker92]
3	2 [Rao]	1 [Dek79]	1
4	1 [Dek79]	1	1

Table 1.3: The smallest ℓ such that ℓ-abelian n^{th} powers are avoidable on an m-letter alphabet.

Some basic facts on ℓ-abelian equivalence are listed in the next lemma.

Lemma 1.44. [KSZ13]

Let $u, v \in A^*$ and $\ell \geq 1$.

- If $u \sim_{ab,\ell} v$, then $u \sim_{ab,\ell'} v$ for all $\ell' \leq \ell$.
- If $u_1 \sim_{ab,\ell} v_1$ and $u_2 \sim_{ab,\ell} v_2$, then $u_1 u_2 \sim_{ab,\ell} v_1 v_2$.

Hence one can define an analogue to factor complexity, with respect to the ℓ-abelian equivalence.

Definition 1.45. The *ℓ-abelian complexity* of an infinite word $\mathbf{w}$ is a function $\mathcal{P}_{\mathbf{w}}^{(\ell)} : \mathbb{N} \to \mathbb{N}$ which maps n to the number of distinct factors of length n occurring in the infinite word $\mathbf{w}$, counted up to ℓ-abelian equivalence. That is, we count ℓ-abelian equivalence classes partitioning the set of factors $\mathrm{Fac}_n(\mathbf{w})$ of length n occurring in $\mathbf{w}$. In particular, for any infinite word $\mathbf{w}$, we have for all $n \geq 0$

$$\mathcal{P}_{\mathbf{w}}^{(1)}(n) \leq \cdots \leq \mathcal{P}_{\mathbf{w}}^{(\ell)}(n) \leq \mathcal{P}_{\mathbf{w}}^{(\ell+1)}(n) \leq \cdots \leq \mathcal{P}_{\mathbf{w}}^{(\infty)}(n).$$

Since we are interested in ℓ-abelian complexity (especially in Chapter 2), it is natural to consider the following operation that permits us to compare factors of length ℓ occurring in an infinite word. Indeed, if two finite words are ℓ-abelian equivalent, it implies that their ℓ-block codings are abelian equivalent, but the converse does not hold (see Example 1.47).

Definition 1.46. Let $\ell \geq 1$. The *ℓ-block coding* of the word $\mathbf{w} = w_0 w_1 w_2 \cdots$ over the alphabet A is the word

$$\mathrm{block}(\mathbf{w}, \ell) = (w_0 \cdots w_{\ell-1})\,(w_1 \cdots w_\ell)\,(w_2 \cdots w_{\ell+1}) \cdots (w_j \cdots w_{j+\ell-1}) \cdots$$

over the alphabet A^ℓ. If $A = \{0, \ldots, r-1\}$, then it is convenient to identify A^ℓ with the set $\{0, \ldots, r^\ell - 1\}$ and each word $w_0 \cdots w_{\ell-1}$ of length ℓ is thus replaced with the integer obtained by reading the word in base r, i.e.,

$$\sum_{i=0}^{\ell-1} w_i\, r^{\ell-1-i}.$$

One can also define similarly the ℓ-block coding of a finite word u of length at least ℓ. The resulting word $\text{block}(u,\ell)$ has length $|u|-\ell+1$.

Example 1.47 (Example 1.42 continued)**.** The 2-block codings of the two 2-abelian equivalent words $u = \mathtt{11010011}$ and $v = \mathtt{11101001}$ are respectively $\mathtt{3212013}$ and $\mathtt{3321201}$, which are abelian equivalent. Consider now the word $\mathtt{0132132}$ which is the 2-block coding of $w = \mathtt{00110110}$. All the 2-block codings are abelian equivalent but $u \not\sim_{ab,2} w \not\sim_{ab,2} v$.

Lemma 1.48. [KSZ13]

Let $\ell \geq 1$. Two finite words u and v of length at least $\ell - 1$ are ℓ-abelian equivalent if and only if they share the same prefix (resp. suffix) of length $\ell - 1$ and the words $\text{block}(u,\ell)$ and $\text{block}(v,\ell)$ are abelian equivalent.

It is well known that the ℓ-block coding of a k-automatic sequence is again a k-automatic sequence [Cob72].

Example 1.49. For the period-doubling word $\mathbf{p}$, the 2-block coding is given by

$$\text{block}(\mathbf{p},2) = \eta^{\omega}(1) = \mathtt{1200121212001200120012121200121 2}\cdots$$

where η is the morphism over $\{\mathtt{0},\mathtt{1},\mathtt{2}\}^*$ defined by $\eta : \mathtt{0} \mapsto \mathtt{12}, \mathtt{1} \mapsto \mathtt{12}, \mathtt{2} \mapsto \mathtt{00}$.

Example 1.50. For the Thue–Morse word $\mathbf{t}$, the 2-block coding is given by

$$\text{block}(\mathbf{t},2) = \nu^{\omega}(1) = \mathtt{132120132012132120121320}\cdots$$

where ν is the morphism over $\{\mathtt{0},\mathtt{1},\mathtt{2},\mathtt{3}\}^*$ defined by $\nu : \mathtt{0} \mapsto \mathtt{12}, \mathtt{1} \mapsto \mathtt{13}$, $\mathtt{2} \mapsto \mathtt{20}, \mathtt{3} \mapsto \mathtt{21}$.

1.5 Sturmian words

Sturmian words form one of the most studied classes of infinite words. For surveys, see [Lot02, Chapter 3] and [PF02, Chapter 6]. It can be defined in several ways. For instance, it can be defined as infinite words having the lowest factor complexity among all aperiodic words.

Definition 1.51. An infinite word $\mathbf{w}$ is *Sturmian* if $\mathcal{P}^{(\infty)}_{\mathbf{w}}(n) = n+1$ for all $n \geq 0$. In particular, Sturmian words are over a binary alphabet as $\mathcal{P}^{(\infty)}_{\mathbf{w}}(1) = 2$.

Example 1.52. The Fibonacci word introduced in Example 1.14

$$\mathbf{f} = \mathtt{0100101001001010010100100101001001}\cdots$$

is a Sturmian word. A short computation shows

$$\begin{aligned} &\mathrm{Fac}_0(\mathbf{f}) = \{\varepsilon\} && \mathcal{P}_{\mathbf{f}}^{(\infty)}(0) = 1 \\ &\mathrm{Fac}_1(\mathbf{f}) = \{\mathtt{0},\mathtt{1}\} && \mathcal{P}_{\mathbf{f}}^{(\infty)}(1) = 2 \\ &\mathrm{Fac}_2(\mathbf{f}) = \{\mathtt{00},\mathtt{01},\mathtt{10}\} && \mathcal{P}_{\mathbf{f}}^{(\infty)}(2) = 3 \\ &\mathrm{Fac}_3(\mathbf{f}) = \{\mathtt{001},\mathtt{010},\mathtt{100},\mathtt{101}\} && \mathcal{P}_{\mathbf{f}}^{(\infty)}(3) = 4. \end{aligned}$$

To show that $\mathcal{P}_{\mathbf{f}}^{(\infty)}(n) = n+1$ for all $n \in \mathbb{N}$, it remains to show that $u \in \mathrm{Fac}_n(\mathbf{f})$ implies $u\mathtt{0} \in \mathrm{Fac}_{n+1}(\mathbf{f}), u\mathtt{1} \notin \mathrm{Fac}_{n+1}(\mathbf{f})$ or $u\mathtt{0} \notin \mathrm{Fac}_{n+1}(\mathbf{f}), u\mathtt{1} \in \mathrm{Fac}_{n+1}(\mathbf{f})$ for all factors u except of one. The proof of this fact can be found in [Lot02, Example 2.1.1].

An equivalent definition of Sturmian words is given in the following theorem. Recall from Section 1.1 that an infinite word $\mathbf{w}$ over the alphabet $\{\mathtt{0},\mathtt{1}\}$ is balanced if for all $u, v \in \mathrm{Fac}(\mathbf{w})$ of same length, we have $||u|_{\mathtt{1}} - |v|_{\mathtt{1}}| \leq 1$.

Theorem 1.53. Coven–Hedlund [CH73, Lot02]

An infinite word $\mathbf{w} \in \{\mathtt{0},\mathtt{1}\}^{\mathbb{N}}$ is Sturmian if and only if it is aperiodic and balanced.

This means that if $\mathbf{w}$ is a Sturmian word, then its abelian complexity satisfies

$$P_{\mathbf{w}}^{(1)}(n) = 2 \quad \text{for all } n \geq 1. \tag{1.1}$$

Theorem 1.54. [KSZ13]

An aperiodic infinite word $\mathbf{w}$ is Sturmian if and only if $\mathcal{P}_{\mathbf{w}}^{(\ell)}(n) = \min(n+1, 2\ell)$ for all $n \geq 1$ and $\ell \in \mathbb{N} \cup \{\infty\}$.

1.5.1 Return words

Another characterization of Sturmian words was established in terms of return words by Vuillon [Vui01]. The classical notion of return words has been used by Durand [Dur98] and was previously introduced by Boshernitzan [Bos85] (see also [DGS76] for the notion of induced transformation in a dynamical context).

Definition 1.55. Let u be a prefix of a uniformly recurrent word $\mathbf{w}$. A non-empty factor w of $\mathbf{w}$ is a *return word* to u if there exists some $i \geq 0$ such that

- $\mathbf{w}[i, i+|w|-1] = w$,
- $\mathbf{w}[i, i+|u|-1] = u = \mathbf{w}[i+|w|, i+|w|+|u|-1]$,
- $\mathbf{w}[i+j, i+j+|u|-1] \neq u$ for all $j \in \{1, \ldots, |w|-1\}$.

Observe that a return word to u may be of smaller length than u (Figure 1.4).

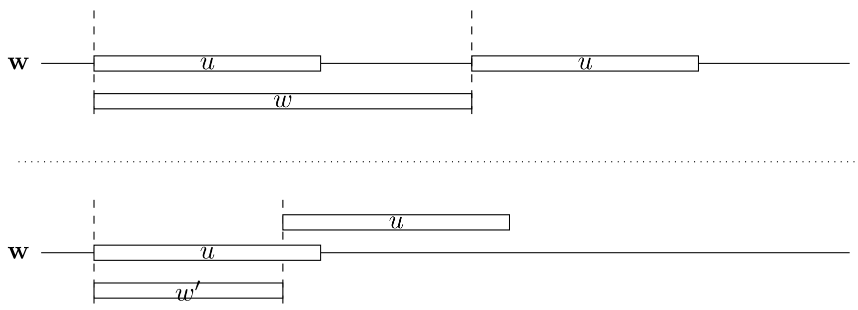


Figure 1.4: Two possible configurations for consecutive occurrences of a factor u in an infinite word $\mathbf{w}$ and their corresponding return words w and w'.

We denote by $\mathcal{R}_{\mathbf{w},u}$ the set of return words to u. Since $\mathbf{w}$ is uniformly recurrent, this set is finite because the length of the longest return word to u is bounded by the maximal distance between two successive occurrences of u. If we order the return words to u with respect to their first occurrence in x, then the corresponding map is

$$\Lambda_{\mathbf{w},u} : \mathcal{R}_{\mathbf{w},u} \to \{\mathtt{1}, \ldots, \#(\mathcal{R}_{\mathbf{w},u})\} =: R_{\mathbf{w},u}.$$

Since $\mathcal{R}_{\mathbf{w},u}$ is a *code* [Dur98], i.e., any element in $\mathcal{R}^*_{\mathbf{w},u}$ has a unique factorization as return words to u, $\mathbf{w}$ can be written in a unique way as $\mathbf{w} = m_0 m_1 m_2 \cdots$. The *derived sequence* $\mathcal{D}_u(\mathbf{w})$ is the infinite word $\Lambda_{\mathbf{w},u}(m_0)\Lambda_{\mathbf{w},u}(m_1)\Lambda_{\mathbf{w},u}(m_2)\cdots$ over $R_{\mathbf{w},u}$.

Proposition 1.56. [Dur98]

Let $\mathbf{w} = \varphi^\omega(a)$ be a fixed point of a primitive morphism φ and let u be a non-empty prefix of $\mathbf{w}$. The derived sequence $\mathcal{D}_u(\mathbf{w})$ is the unique fixed point of the unique prolongable morphism σ_u satisfying $\Theta_{\mathbf{w},u}\sigma_u = \sigma\Theta_{\mathbf{w},u}$ where $\Theta_{\mathbf{w},u}$ denotes the inverse map of $\Lambda_{\mathbf{w},u}$.

Example 1.57. Consider the Fibonacci word $\mathbf{f}$ and its prefix $u = \mathtt{010}$. We mark the beginning of occurrences of u by vertical lines in the following prefix of $\mathbf{f}$:

$$|\mathtt{010}|\mathtt{01}|\mathtt{010}|\mathtt{010}|\mathtt{01}|\mathtt{010}|\mathtt{01}|\mathtt{010}|\mathtt{010}|\mathtt{01}|\mathtt{010}|\mathtt{010}|\mathtt{01}|\mathtt{010}\cdots.$$

Hence, the set of return words to u is the set $\mathcal{R}_{\mathbf{f},u} = \{\mathtt{010}, \mathtt{01}\}$ where the words are written in the order of their first occurrences in $\mathbf{f}$. So, $R_{\mathbf{f},u} = \{1, 2\}$ and the map $\Lambda_{\mathbf{f},u}$ is defined by $\mathtt{010} \mapsto \mathtt{1}, \mathtt{01} \mapsto \mathtt{2}$. The derived sequence is then

$$\mathcal{D}_u(\mathbf{f}) = \mathtt{12112121121121211212112112121121121121}\cdots.$$

Using $\Theta_{\mathbf{f},u}, \Lambda_{\mathbf{f},u}$ and $\varphi_{\mathbf{f}}$, we compute the morphism $\sigma_{\mathtt{010}}$:

$$\mathtt{1} \xrightarrow{\Theta_{\mathbf{w},u}} \mathtt{0100} \xrightarrow{\varphi_{\mathbf{f}}} \mathtt{01001010} \xrightarrow{\Lambda_{\mathbf{w},u}} \mathtt{12},$$
$$\mathtt{2} \xrightarrow{\Theta_{\mathbf{w},u}} \mathtt{010} \xrightarrow{\varphi_{\mathbf{f}}} \mathtt{010010} \xrightarrow{\Lambda_{\mathbf{w},u}} \mathtt{1}.$$

We find that $\sigma_{\mathtt{010}} : \mathtt{1} \mapsto \mathtt{12}, \mathtt{2} \mapsto \mathtt{12}$. So the derived sequence $\mathcal{D}_u(\mathbf{f})$ is simply the Fibonacci word up to a relabelling of the letters.

Durand [Dur98] used the notion of derived sequences to characterize primitive substitutive word. Recall from Section 1.2 that a primitive substitutive word is an infinite word of the form $\varphi^\omega(a)$ where the morphism φ is such that for some positive integer k, $|\varphi^k(b)|_c \geq 1$ for all letters $a, b \in A$.

Theorem 1.58. [Dur98]

A word is primitive substitutive if and only if the number of its different derived sequences is finite.

Vuillon [Vui01] considered the number of return words with respect to each factor separately to obtain a characterization of Sturmian words.

Theorem 1.59. [Vui01]

A recurrent infinite word is Sturmian if and only if each of its factor has two return words.

There also exists a simple characterization of periodicity via return words.

Proposition 1.60. [Vui01]

A recurrent infinite word is periodic if and only if there exists a factor having exactly one return word.

1.5.2 Rotation words

Many of our results in Chapter 3 on Sturmian words rely on the definition of Sturmian words in terms of *rotation words*. Rotation word can be obtained by coding the orbit $(R_\alpha^n(\rho))_{n\geq 0}$ of a point ρ on a circle under a rotation R_α by an angle α when the circle is partitioned in a suitable way into complementary intervals. See for instance [Rig14].

Example 1.61. Take three distinct points $\lambda_1, \lambda_2, \lambda_3$ on the circle identified with the interval $[0, 1)$ and set $\lambda_0 = 0$. In Figure 1.5, we choose

$$\lambda_1 = 1.8/(2Pi) \approx 0.286,\ \lambda_2 = 3.67/(2\pi) \approx 0.584,\ \lambda_3 = 4.82/(2\pi) \approx 0.767.$$

Then the coding of the orbit of a point $\rho \in [0, 1)$ under a rotation of angle α is the word $\mathbf{w}_\rho = w_0 w_1 \cdots$ where $w_i = j$ if and only if $R_\alpha^i(\rho) \in [\lambda_j, \lambda_{j+1})$. If we set $\alpha = 3/(8\pi) \approx 0.119$ and $\rho = 0.08$ as in Figure 1.5, we obtain

$$\mathbf{w}_\rho = \mathtt{0011123300011223} \cdots .$$

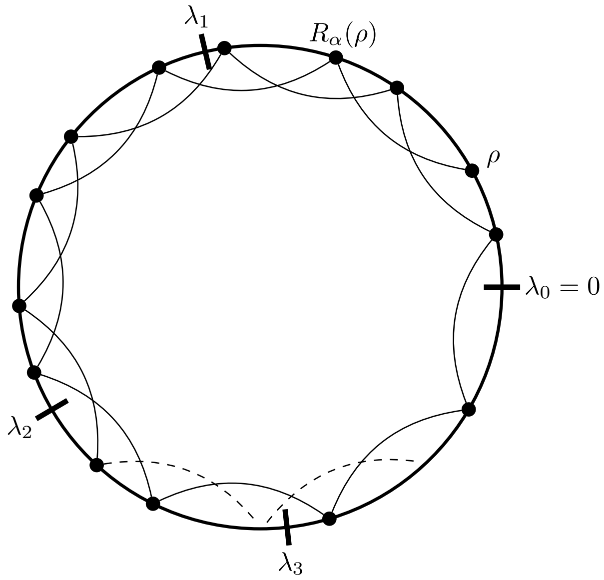


Figure 1.5: The first points of an orbit under a rotation R_α of angle α.

Let $\mathcal{C}$ be the one-dimensional torus $\mathbb{R}/\mathbb{Z}$ identified with the interval $[0, 1)$. As usual, we denote by $\{x\}$ the fractional part of x. The *rotation* R_α defined for a real number α is the map

$$R_\alpha : \mathcal{C} \to \mathcal{C}, x \mapsto \{x + \alpha\}.$$

By an *interval* (resp. *half-interval*) of $\mathcal{C}$ we mean a set of points that is an image of an interval (resp. half-interval) of $\mathbb{R}$ under operation $\{\cdot\}$. For instance, if $0 \leq b < a < 1$, then $[a, 1) \cup [0, b)$ is denoted by $[a, b)$.

In this setting, if the angle α is a rational number, then it would only produce periodic orbits and so periodic words [Lot02]. Hence, we only consider irrational angles in the sequel. Recall a set S is *dense* in a topological space X if its closure $\overline{S}$ is equal to the whole space X.

Theorem 1.62. Kronecker[AS03a]

If a number α is irrational, then the set of points $\{R^i_\alpha(\rho) \mid i \in \mathbb{N}\}$ is dense in $\mathcal{C}$ for all initial points $\rho \in \mathcal{C}$.

Sturmian words are particular rotation words with an angle α irrational, a starting point ρ real and only two intervals partitioning $\mathcal{C}$. Without loss of generality we can assume $0 \leq \alpha, \rho < 1$.

Proposition 1.63. [AS03a, Lot02]

Let $\alpha \in [0,1)$ be irrational and $\rho \in [0,1)$ be real. An infinite word $\mathbf{w}$ is Sturmian if and only if any letter w_i of $\mathbf{w}$ satisfies

$$w_i = \begin{cases} \mathtt{0} & \text{if } R_\alpha^i(\rho) \in I_0 \\ \mathtt{1} & \text{if } R_\alpha^i(\rho) \in I_1 \end{cases}$$

where $I_0 = [0, 1-\alpha)$ and $I_1 = [1-\alpha, 1)$ (respectively $I_0 = (0, 1-\alpha]$ and $I_1 = (1-\alpha, 1]$). In which case, we write $\mathbf{w} = St(\alpha, \rho)$ (resp. $\mathbf{w} = St'(\alpha, \rho)$).

The irrational α is called the *slope* of the Sturmian word $St(\alpha, \rho)$ and the initial point ρ is its *intercept*. If $\rho = 0$, then

$$St(\alpha, 0) = \mathtt{0}\mathbf{c}_\alpha \text{ and } St'(\alpha, 0) = \mathtt{1}\mathbf{c}_\alpha$$

and $\mathbf{c}_\alpha$ is said to be the *characteristic Sturmian word of slope* α [Lot02]. If $\mathbf{w} = St(\alpha, 0)$, we say that $\mathbf{w}$ is a Sturmian word *with null intercept.*

Example 1.64. The Fibonacci word is a Sturmian word with parameters α and ρ both equal to $1/\phi^2 = 2 - \phi \approx 0.38197$ where $\phi = (1+\sqrt{5})/2$ is the Golden mean:

$$\mathbf{f} = St(1/\phi^2, 1/\phi^2) = \mathtt{0100101001001010010100100101001001} \cdots .$$

Then the characteristic word of $St(1/\phi^2, 0)$ is the Fibonacci word.

Chapter 2

Regularity of ℓ-abelian complexity functions

This chapter, based on a joint work with my co-advisor Rigo and two postdoctoral fellows Parreau and Rowland, is about some structural properties of integer sequences that occur naturally in combinatorics on words. We prove that a sequence satisfying a certain symmetry property is 2-regular in the sense of Allouche and Shallit. We apply this theorem to show that both the period-doubling word and the Thue–Morse word have 2-abelian complexity sequences which are 2-regular. The computations and arguments leading to these results permit us to exhibit some similarities between the two cases and a quite general scheme that we hope can be used again to prove additional regularity results. Indeed, we conjecture that *any k-automatic sequence has an ℓ-abelian complexity function that is k-regular.*

Contents

Recently there has been a renewal of interest in abelian notions arising in combinatorics on words. For instance, avoiding abelian or ℓ-abelian repetitions (see Table 1.3), counting the number of abelian bordered words (see for instance [RRS13, CHPZ14]), and so on.

A year ago, Madill and Rampersad considered the abelian complexity of the ordinary paperfolding word [MR13]. There are several equivalent definitions of the ordinary paperfolding word. For instance, consider a morphism $\theta : \{a, b, c, d\}^* \to \{a, b, c, d\}^*$ and a coding $\eta : \{a, b, c, d\}^* \to \{\mathtt{0}, \mathtt{1}\}^*$ defined by

$$\theta : \begin{cases} a \mapsto ab \\ b \mapsto ac \\ c \mapsto dc \\ d \mapsto db \end{cases} \text{ and } \eta : \begin{cases} a \mapsto \mathtt{0} \\ b \mapsto \mathtt{0} \\ c \mapsto \mathtt{1} \\ d \mapsto \mathtt{1}. \end{cases}$$

The *paperfolding word* is the coding of the fixed point of θ starting with a

$$\eta(\theta^\omega(a)) = 001001100011011000100111001101100010011 0 \cdots$$

This sequence is another example of 2-automatic sequences since it is generated by iteration of a 2-uniform morphism and an application of a coding. Madill and Rampersad were the first to precisely compute the abelian complexity of an infinite word in the case where this complexity grows unboundedly large. They showed that the abelian complexity of the ordinary paperfolding word is 2-regular by proving a finite list of recurrence relations that determine the complexity.

Theorem 2.1. [MR13]

The abelian complexity $\mathcal{P}^{(1)}_{\mathbf{w}}$ of the ordinary paperfolding word $\mathbf{w}$ satisfies, for all $n \geq 0$,

$$\begin{aligned} \mathcal{P}^{(1)}_{\mathbf{w}}(4n) &= \mathcal{P}^{(1)}_{\mathbf{w}}(2n) \\ \mathcal{P}^{(1)}_{\mathbf{w}}(4n+2) &= \mathcal{P}^{(1)}_{\mathbf{w}}(2n+1)+1 \\ \mathcal{P}^{(1)}_{\mathbf{w}}(16n+1) &= \mathcal{P}^{(1)}_{\mathbf{w}}(8n+1) \\ \mathcal{P}^{(1)}_{\mathbf{w}}(16n+\{3,7,9,13\}) &= \mathcal{P}^{(1)}_{\mathbf{w}}(2n+1)+2 \\ \mathcal{P}^{(1)}_{\mathbf{w}}(16n+5) &= \mathcal{P}^{(1)}_{\mathbf{w}}(4n+1)+2 \\ \mathcal{P}^{(1)}_{\mathbf{w}}(16n+11) &= \mathcal{P}^{(1)}_{\mathbf{w}}(4n+3)+2 \\ \mathcal{P}^{(1)}_{\mathbf{w}}(16n+15) &= \mathcal{P}^{(1)}_{\mathbf{w}}(2n+2)+1. \end{aligned}$$

Since then, other (unbounded) abelian complexities were studied. For instance, the (ℓ-) abelian complexities of the Thue–Morse word $\mathbf{t}$, introduced in Example 1.11, and the period-doubling word $\mathbf{p}$, introduced in Example 1.16, were considered. In [KSZ], the authors studied the asymptotic behaviour of $\mathcal{P}^{(\ell)}_{\mathbf{t}}(n)$ and also derived some recurrence relations showing that the abelian complexity $\mathcal{P}^{(1)}_{\mathbf{p}}(n)_{n\geq 0}$ of the period-doubling word $\mathbf{p}$ is 2-regular.

Theorem 2.2. [KSZ, BSCRF14]

The abelian complexity $\mathcal{P}_{\mathbf{p}}^{(1)}$ of the period-doubling word $\mathbf{p}$ satisfies, for all $n \geq 1$,

$$\begin{aligned}
\mathcal{P}_{\mathbf{p}}^{(1)}(2n) &= \mathcal{P}_{\mathbf{p}}^{(1)}(n) \\
\mathcal{P}_{\mathbf{p}}^{(1)}(4n-1) &= \mathcal{P}_{\mathbf{p}}^{(1)}(n) + 1 \\
\mathcal{P}_{\mathbf{p}}^{(1)}(4n+1) &= \mathcal{P}_{\mathbf{p}}^{(1)}(n) + 1.
\end{aligned}$$

In [BSCRF14], Blanchet-Sadri et al. studied the abelian complexity of the fixed point $\mathbf{v}$ of the non-uniform morphism $\mathtt{0} \mapsto \mathtt{012}, \mathtt{1} \mapsto \mathtt{02}, \mathtt{2} \mapsto \mathtt{1}$ and they obtain results similar to those discussed in this chapter. Even though they are not directly interested in the k-regularity of $\mathcal{P}_{\mathbf{v}}^{(1)}(n)_{n\geq 0}$, they derive some recurrence relations. From these relations, following the approach described in this chapter, one can possibly prove some regularity results. In particular, the result of replacing in $\mathbf{v}$ all 2's by 0's leads back to the period-doubling word. Hence, Blanchet-Sadri et al. also proved the relations about the abelian complexity of $\mathbf{p}$ given in Theorem 2.2.

Let us now describe the content and organization of this chapter, which is based on a joint work with my co-advisor Rigo and two post-doctoral fellows at the University of Liège, Parreau and Rowland [PRRV].

Using matrices, we compute in Section 2.1 the first 65538 values of the 2-abelian complexity of the Thue–Morse word introduced in Example 1.11. From these values, we conjecture the 2-regularity of this sequence and some relations. Recently, after hearing a talk I gave during the *Representing Streams II* meeting in January 2014, Greinecker proved the recurrence relations needed to prove the 2-regularity of this sequence [Gre]. Hopefully, the two approaches are complementary: we prove the 2-regularity without exhibiting the explicit recurrence relations.

In Section 2.2, we prove the 2-regularity of a large family of sequences satisfying a recurrence relation with a parameter c and 2^{ℓ_0} initial conditions. The form of the recurrence implies that sequences in this family exhibit a reflection symmetry in the values taken over each interval $[2^\ell, 2^{\ell+1})$ for $\ell \geq \ell_0$. For the special case of the Thue–Morse word, a similar property is shown in [Gre]. Computer experiments suggest that many 2-abelian complexity functions satisfy such a reflection property. These functions are given in Appendix A.

In Section 2.3, we study the abelian complexity of the 2-block coding $\mathbf{y} = \mathrm{block}(\mathbf{t}, 2)$ of the Thue–Morse word $\mathbf{t}$, introduced in Example 1.50. We consider the sum of occurrences of $\mathtt{1}$ and $\mathtt{2}$ in each factor of length n in $\mathbf{y}$ and we define $\Delta_{12}(n)$ to be the difference between the maximal sum and minimal sum for factors of fixed length n in $\mathbf{y}$. It turns out that $\Delta_{12}(n) + 1 = \mathcal{P}_{\mathbf{p}}^{(1)}(n)$ and our results can thus be related to [BSCRF14] and [KSZ]. We prove that $\Delta_{12}(n)_{n\geq 0}$ and $\mathcal{P}_{\mathbf{y}}^{(1)}(n)_{n\geq 0}$ are 2-regular. We show that the 2-regularity of $\mathcal{P}_{\mathbf{t}}^{(2)}(n)$ follows from the 2-regularity of $\Delta_{12}(n)_{n\geq 0}$ and $\mathcal{P}_{\mathbf{y}}^{(1)}(n)_{n\geq 0}$.

Section 2.4 shares some similarities with Section 2.3. The reader will see that the strategy used to prove the 2-regularity of $\mathcal{P}_{\mathbf{t}}^{(2)}(n)$ can also be applied to the 2-abelian complexity of the period-doubling word. Nevertheless, some differences do not permit us to treat the two cases within a completely unified framework.

In Section 2.4, we study the abelian complexity of the 2-block coding $\mathbf{x} = \mathrm{block}(\mathbf{p}, 2)$ of the period-doubling word $\mathbf{p}$, introduced in Example 1.49. In particular, we consider the difference $\Delta_0(n)$ between the maximal and minimal numbers of 0's occurring in factors of length n in block$(\mathbf{p}, 2)$. We prove that the sequences $\Delta_0(n)_{n\geq 0}$ and $\mathcal{P}^{(1)}_{\mathbf{x}}(n)_{n\geq 0}$ are 2-regular. Then, we study the 2-abelian complexity of $\mathbf{p}$. We show that the 2-regularity of $\Delta_0(n)_{n\geq 0}$ and $\mathcal{P}^{(1)}_{\mathbf{x}}(n)_{n\geq 0}$ implies the 2-regularity of $\mathcal{P}^{(2)}_{\mathbf{p}}(n)$.

Finally, in Section 2.5 we suggest a direction for future work.

2.1 Origin of the result on the 2-abelian complexity of Thue–Morse word t

The first few terms of the 2-abelian complexity $\mathcal{P}^{(2)}_{\mathbf{t}}(n)_{n\geq 0}$ of the Thue–Morse word $\mathbf{t}$ are depicted in Figure 2.1 and given by

$$1, 2, 4, 6, 8, 6, 8, 10, 8, 6, 8, 8, 10, 10, 10, 8, 8, 6, 8, 10, 10, 8, 10, 12, 12, 10, 12, 12, \ldots$$

Naturally, one can ask whether the 2-abelian complexity of $\mathbf{t}$ is bounded or whether it

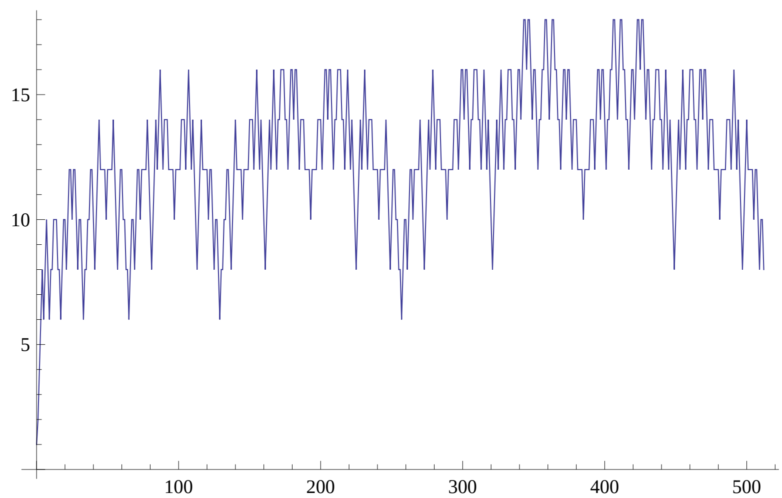


Figure 2.1: The 2-abelian complexity of $\mathbf{t}$ on the interval $[0, 512]$.

is regular. The first question was answered independently by Berthé and Delecroix [BD] and Karhumäki et al. [KSZ]. Following the result [Ada03, Corollary 15], Berthé and Delecroix [BD] showed that the function

$$n \mapsto \max_{u\in \mathrm{Fac}_2(\mathbf{t})} \max_{w,w'\in \mathrm{Fac}_n(\mathbf{t})} ||w|_u - |w'|_u|$$

is not bounded. In other words, they showed that the 2-abelian complexity of $\mathbf{t}$ is not bounded. Karhumäki et al. [KSZ] studied the behaviour of the 2-abelian complexity of $\mathbf{t}$ and proved that for $n \geq 1$ and $m \geq 0$,

$$\mathcal{P}^{(2)}_{\mathbf{t}}(n) = O(\log n),\ \mathcal{P}^{(2)}_{\mathbf{t}}((2\cdot 4^m + 4)/3) = \Theta(m) \text{ and } \mathcal{P}^{(2)}_{\mathbf{t}}(2^m + 1) \leq 8$$

which leads to unbounded 2-abelian complexity.

The second question, whether the 2-abelian complexity is regular, was solved independently by Greinecker [Gre] and ourselves [PRRV]. My co-advisor Rigo and I conjectured recurrence relations that imply the 2-regularity of the complexity [RV12]. Greinecker [Gre] proved the conjecture relations, while we [PRRV] proved the 2-regularity without exhibiting the explicit recurrence relations. To make this conjecture, we needed to compute "quickly" a long enough prefix of the 2-abelian complexity. First, we present the method used to compute such a prefix. Then, we show how the conjecture was established.

2.1.1 We compute the 2-abelian complexity of t using matrix product

We can easily compute the first few terms of the 2-abelian complexity $\mathcal{P}_{\mathbf{t}}^{(2)}(n)_{n\geq 0}$ of the Thue–Morse word, using matrix product. The `Mathematica` code corresponding to this computation is given in Appendix A. Note that this computation is faster than a more naive approach (see Remark A.1).

First, to any word $u = u_1u_2\cdots u_{n-1}u_n$, we associate a vector of $\mathbb{N}^{10}$:

$$\Psi_2(u) = \begin{pmatrix} |u_1|_{\mathtt{0}} \\ |u_1|_{\mathtt{1}} \\ |u|_{\mathtt{00}} \\ |u|_{\mathtt{01}} \\ |u|_{\mathtt{10}} \\ |u|_{\mathtt{11}} \\ |u_{n-1}u_n|_{\mathtt{00}} \\ |u_{n-1}u_n|_{\mathtt{01}} \\ |u_{n-1}u_n|_{\mathtt{10}} \\ |u_{n-1}u_n|_{\mathtt{11}} \end{pmatrix}.$$

For example, $\Psi_2(11101) = (0,1,0,1,1,2,0,1,0,0)$. Following the equivalent definition of ℓ-abelian equivalence given in Lemma 1.43, two words u and v of the same length $n \geq 2$ are 2-abelian equivalent if and only if

- the first six entries of $\Psi_2(u)$ and $\Psi_2(v)$ coincide,
- $[\Psi_2(u)]_7 + [\Psi_2(u)]_9 = [\Psi_2(v)]_7 + [\Psi_2(v)]_9$,
- $[\Psi_2(u)]_8 + [\Psi_2(u)]_{10} = [\Psi_2(v)]_8 + [\Psi_2(v)]_{10}$.

In this case, we say that their vectors are *similar* and denote it by $\Psi_2(u) \sim \Psi_2(v)$. Observe that the last two items are implied by the first one by Lemma 1.43. Indeed, the last two items translate that the last letters of u and v must coincide, since u ends with `1` if and only if u ends either with `01` or `11`, i.e., either $[\Psi_2(u)]_7 = 1$ or $[\Psi_2(u)]_9 = 1$.

Consider all factors of length 3 in the Thue–Morse word: `001`, `010`, `011`, `100`, `101`, `110`. They are non 2-abelian equivalent, since their vectors are not similar (Table 2.1).

Second, we obtain the vector $\Psi_2(v)$ of a factor v of $\mathbf{t}$ from the vector of a shorter factor u occurring before v. Assume that v is of length $2n-1$ and occurs at an even index $2j$. Then v is the word $\sigma(u)$ without its last letter, where u denotes the length-n factor occurring at index j. Similarly, if v occurs at index $2j+1$, then v is $\sigma(u)$ without its first letter (Figure 2.2). We define a matrix for each case, M_e and M_o, so that

- $\Psi_2(v) = M_e\Psi_2(u)$ if v occurs at an even index with $|v| = 2n-1$,

$u = u_1 \cdots u_n$	001	010	011	100	101	110
$\lvert u_1 \rvert_0$	1	1	1	0	0	0
$\lvert u_1 \rvert_1$	0	0	0	1	1	1
$\lvert u \rvert_{00}$	1	0	0	1	0	0
$\lvert u \rvert_{01}$	1	1	1	0	1	0
$\lvert u \rvert_{10}$	0	1	0	1	1	1
$\lvert u \rvert_{11}$	0	0	1	0	0	1
$\lvert u_{n-1}u_n \rvert_{00}$	0	0	0	1	0	0
$\lvert u_{n-1}u_n \rvert_{01}$	1	0	0	0	1	0
$\lvert u_{n-1}u_n \rvert_{10}$	0	1	0	0	0	1
$\lvert u_{n-1}u_n \rvert_{11}$	0	0	1	0	0	0

Table 2.1: Vectors of length-3 factors of $\mathbf{t}$.

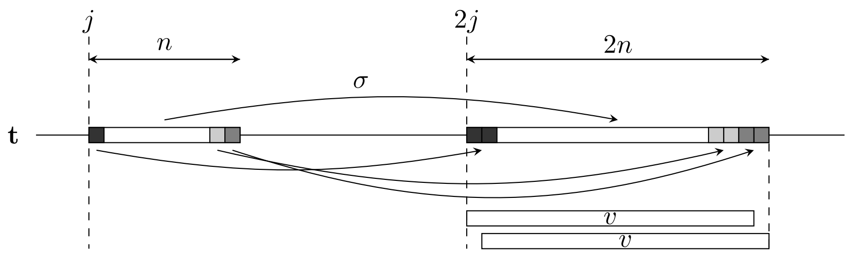


Figure 2.2: From a factor of length n to a factor v of length $2n - 1$.

- $\Psi_2(v) = M_o \Psi_2(u)$ if v occurs at an odd index with $|v| = 2n - 1$.

Assume now that v is of length $2n - 2$. If v occurs at index $2j$, then it corresponds to $\sigma(u)$ without its last two letters. If v occurs at index $2j + 1$, then v is $\sigma(u)$ without its first and last letters (Figure 2.3). Again, we define a matrix for each case, N_e and N_o, so that

- $\Psi_2(v) = N_e \Psi_2(u)$ if v occurs at an even index with $|v| = 2n - 2$,

- $\Psi_2(v) = N_o \Psi_2(u)$ if v occurs at an odd index with $|v| = 2n - 2$.

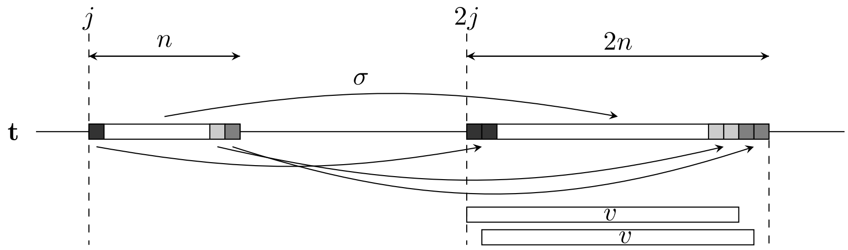


Figure 2.3: From a factor of length n to a factor v of length $2n - 2$.

A similar structure appears in all these four matrices:

$$M_e := \left(\begin{array}{cc|cccc|cccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 1 & -1 & 0 & -1 & 0 \\
0 & 1 & 1 & 1 & 0 & 1 & 0 & -1 & 0 & -1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0
\end{array}\right),$$

$$M_o := \left(\begin{array}{cc|cccc|cccc}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right),$$

$$N_e := \left(\begin{array}{cc|cccc|cccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\
1 & 0 & 1 & 0 & 1 & 1 & -1 & 0 & -1 & -1 \\
0 & 1 & 1 & 1 & 0 & 1 & -1 & -1 & 0 & -1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ \hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right),$$

$$\text{and } N_o := \left(\begin{array}{cc|cccc|cccc}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 & -1 & 0 & -1 & 0 \\
0 & 0 & 1 & 1 & 0 & 1 & 0 & -1 & 0 & -1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0
\end{array}\right).$$

Indeed, we can view them as matrices with fives blocks that are possibly distinct from 0:

$$\left(\begin{array}{c|c|c}
B_1 & 0 & 0 \\ \hline
C & B_2 & D \\ \hline
0 & 0 & B_3
\end{array}\right).$$

On the diagonal, the first block B_1 takes care of the first letter. Similarly, the third block B_3 takes care of the last two letters. The second block B_2 takes care of the number of length-2 factors occurring in $\sigma(u)$ without its first letter. Since $\mathtt{0}^{-1}\sigma(00) = \mathtt{0}^{-1}\mathtt{0101} = \mathtt{101}$ contains one $\mathtt{10}$ and one $\mathtt{01}$, the first column of B_2 is $(0, 1, 1, 0)$. Counting the length-2 factors occurring in the words $\mathtt{0}^{-1}\sigma(01), \mathtt{1}^{-1}\sigma(10), \mathtt{1}^{-1}\sigma(11)$, we obtain respectively the last three columns of B_2. For example, if u is the factor $\mathtt{011001011}$ occurring at index 10 in $\mathbf{t}$, then one can check that $\mathtt{0}^{-1}\sigma(u) = \mathtt{11010010110011010}$ has two occurrences $\mathtt{00}$, five occurrences $\mathtt{01}$, six occurrences $\mathtt{10}$ and three occurrences $\mathtt{11}$, and

$$B_2\left(|u|_{\mathtt{00}}, |u|_{\mathtt{01}}, |u|_{\mathtt{10}}, |u|_{\mathtt{11}}\right) = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \\ 6 \\ 3 \end{pmatrix}.$$

The block C is different from 0 only in M_e and N_e, i.e., for factors occurring at even indices. This block takes into account the first length-2 factor coming from the image of the first letter of u. Finally, the block D deals with the length-2 factors coming from the image of the last letter of u. It is distinct from 0 for M_e, N_e, N_o because these matrices correspond to the factor $\sigma(u)$ without its last letter or its two last letters.

These four matrices allow us to provide the set of finite factors of $\mathbf{t}$ with the structure of a tree (Figure 2.4). Let $S_3 = \{\Psi_2(u) \mid u \in \mathrm{Fac}(\mathbf{t}), |u| = 3\}$ and, for all $\ell \geq 2$ and $0 \leq r < 2^\ell$,

$$S_{2^\ell+r} = \begin{cases} \{N_e y \mid y \in S_{2^{\ell-1}+\frac{r}{2}+1}\} \cup \{N_o y \mid y \in S_{2^{\ell-1}+\frac{r}{2}+1}\} & \text{if } r \text{ is even} \\ \{M_e y \mid y \in S_{2^{\ell-1}+\frac{r+1}{2}}\} \cup \{M_o y \mid y \in S_{2^{\ell-1}+\frac{r+1}{2}}\} & \text{if } r \text{ is odd.} \end{cases}$$

Then, we have $\mathcal{P}^{(2)}_{\mathbf{t}}(n) = \#(S_n/{\sim})$ for $n \geq 3$. Therefore, we are able to easily compute[1] the first values of the $\mathcal{P}^{(2)}_{\mathbf{t}}$.

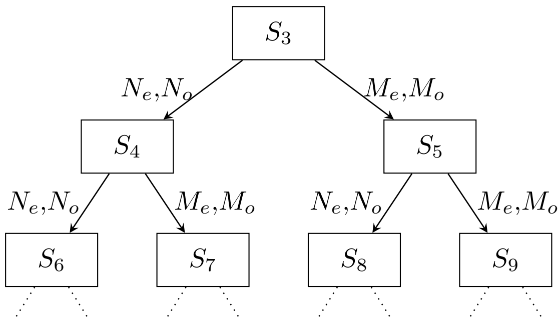


Figure 2.4: Tree structure of the set of finite factors of $\mathbf{t}$.

Lemma 2.3.

Let n be a positive integer and u, v be two factors of length n of the Thue–Morse word $\mathbf{t}$.

1. We have $M_e\Psi_2(u) \not\sim M_o\Psi_2(u)$.
2. If $\Psi_2(u) \sim \Psi_2(v)$, then $M_e\Psi_2(u) \sim M_e\Psi_2(v)$ and $M_o\Psi_2(u) \sim M_o\Psi_2(v)$. But the conserve does not hold in general.
3. We have

$$M_e\Psi_2(u) \sim M_e\Psi_2(v) \Leftrightarrow M_o\Psi_2(u) \sim M_o\Psi_2(v)$$
$$M_e\Psi_2(u) \sim M_o\Psi_2(v) \Leftrightarrow M_o\Psi_2(u) \sim M_e\Psi_2(v)$$

4. The factor u starts and ends with the same letter if and only if $M_e\Psi_2(u) \sim M_o\Psi_2(\overline{u})$, which is equivalent to $M_o\Psi_2(u) \sim M_e\Psi_2(\overline{u})$.

Proof. Let n be a positive integer and $u = u_0 \dots u_{n-1}, v = v_0 \dots v_{n-1}$ be two factors of length n of the Thue–Morse word $\mathbf{t}$ with $u_i, v_i \in \{\mathtt{0}, \mathtt{1}\}$ for $0 \geq i \geq n-1$. Let us write $[\Psi_2(u)]_i = y_i$ and $[\Psi_2(v)]_i = z_i$ for $1 \leq i \leq 10$.

A straightforward computation shows the first assertion, $M_e\Psi_2(u) \not\sim M_o\Psi_2(u)$, since $y_1 = |u_0|_{\mathtt{0}} \neq |u_0|_{\mathtt{1}} = y_2$.

[1]The `Mathematica` code is available in Appendix A

For the second assertion, assume $\Psi_2(u) \sim \Psi_2(v)$, i.e., $y_i = z_i$ for $1 \leq i \leq 6$, $y_7 + y_9 = z_7 + z_9$ and $y_8 + y_{10} = z_8 + z_{10}$. In this case, we have

$M_e\Psi_2(u)$	$M_o\Psi_2(u)$	$M_e\Psi_2(u)$ $- M_e\Psi_2(v)$	$M_o\Psi_2(u)$ $- M_o\Psi_2(v)$
y_1	y_2	0	0
y_2	y_1	0	0
y_5	y_4	0	0
$y_1 + y_3 + y_5 + y_6 - y_7 - y_9$	$y_3 + y_5 + y_6$	0	0
$-y_{10} + y_2 + y_3 + y_4 + y_6 - y_8$	$y_3 + y_4 + y_6$	0	0
y_4	y_5	0	0
y_9	0	$y_9 - z_9$	0
y_{10}	$y_7 + y_9$	$y_{10} - z_{10}$	0
y_7	$y_{10} + y_8$	$y_7 - z_7$	0
y_8	0	$y_8 - z_8$	0

and clearly the two last columns of this table are similar to the null vector $0 \in \mathbb{N}^{10}$. To show that the converse does not hold in general, consider the factors $u = \mathtt{01001}$ and $v = \mathtt{01011}$ of $\mathbf{t}$. They are not 2-abelian equivalent: $\Psi_2(u) \not\sim \Psi_2(v)$. But one can check that $M_e\Psi_2(u) \sim M_e\Psi_2(v)$ and $M_o\Psi_2(u) \sim M_o\Psi_2(v)$. In other words, one can check that the associated factors are such that

$$\sigma(u)\mathtt{0}^{-1} \sim_{ab,2} \sigma(v)\mathtt{0}^{-1} \text{ and } \mathtt{0}^{-1}\sigma(u) \sim_{ab,2} \mathtt{0}^{-1}\sigma(v).$$

The third assertion follows from computation:

$$M_e\Psi_2(u) \sim M_e\Psi_2(v) \Leftrightarrow \begin{cases} y_i = z_i & \text{for } i = 1,2,4,5 \\ y_3 + y_6 = z_3 + z_6 \\ y_7 + y_9 = z_7 + z_9 \\ y_8 + y_{10} = z_8 + z_{10} \end{cases}$$
$$\Leftrightarrow M_o\Psi_2(u) \sim M_o\Psi_2(v).$$

Similarly,

$$M_e\Psi_2(u) \sim M_o\Psi_2(v) \Leftrightarrow \begin{cases} y_1 = z_2 \\ y_2 = z_1 \\ y_1 + y_3 + y_6 - y_7 - y_9 = z_3 + z_6 \\ y_2 + y_3 + y_6 - y_8 - y_{10} = z_3 + z_6 \\ y_4 = z_4 \\ y_5 = z_5 \\ y_7 + y_9 = z_8 + z_{10} \\ y_8 + y_{10} = z_7 + z_9 \end{cases}$$
$$\Leftrightarrow M_o\Psi_2(u) \sim M_e\Psi_2(v).$$

For the last assertion, the factor u starts and ends with the same letter if and only if

$$y_1 = y_7 + y_9, y_2 = y_8 + y_{10}, y_4 = y_5$$

which is equivalent to $M_e\Psi_2(u) \sim M_o\Psi_2(\overline{u})$. The last equivalence of the assertion follows from the previous assertion where we set $v = \overline{u}$. $\square$

Remark 2.4. We can illustrate the third assertion of the previous lemma by a graph[2] where the vertices are equivalence classes of vectors for the relation $\sim$ and arrows correspond to matrix products. Consider two factors u and v of the Thue–Morse word that are not 2-abelian

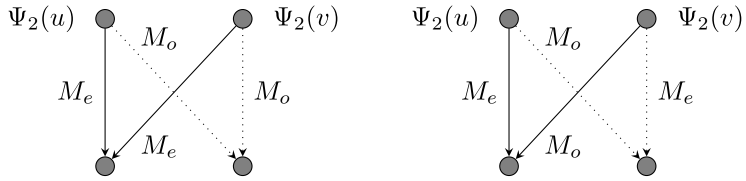
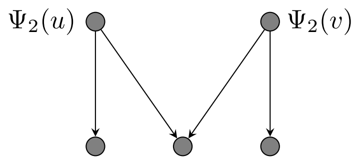


Figure 2.5: Two possible configurations of equivalence classes of vectors.

equivalent. In Figure 2.5, the configuration with the plain edges implies the configuration with the dotted edges, and vice versa. In particular, it means that the following configuration never happens:



or, in other words, two distinct classes cannot split into three classes after multiplying by M_e and M_o.

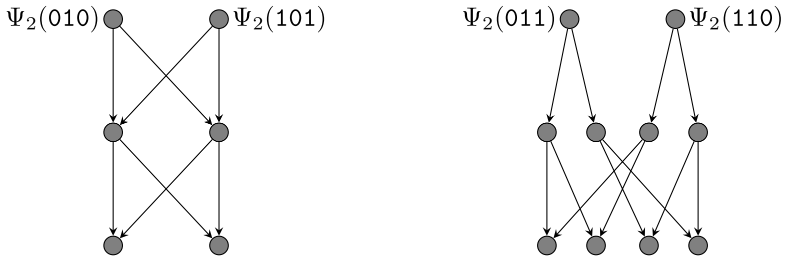
Example 2.5. Using the same notation as in the previous remark, we consider the effect of two consecutive matrix products with M_e or M_o applied to $\Psi_2(u)$ and $\Psi_2(\overline{u})$. For instance, consider the factor $u = \mathtt{010}$ which begins and ends with the same letter. The vectors $\Psi_2(\mathtt{010})$ and $\Psi_2(\overline{\mathtt{010}})$ are not similar. If we apply once or twice a matrix product with M_e and M_o, we obtain the same number of equivalence classes (Figure 2.6(a)) which is 2.

Now consider the factor $u = \mathtt{011}$ which begins and ends with different letters. Again $\Psi_2(\mathtt{011})$ and $\Psi_2(\overline{\mathtt{100}})$ are not similar and applying once or twice a matrix product with M_e and M_o yields the same number of equivalence classes (Figure 2.6(b)) which is 4.

Let n be the length of $u \in \{\mathtt{010}, \mathtt{011}\}$. Then, factors of length $4n + 1$ that are obtained from u (applying the Thue–Morse morphism σ twice and deleting a letter each time) form the same number of 2-abelian equivalence classes as factors of length $2n+1$ that are obtained from u (applying the σ once and deleting a letter).

The result obtained for the factors 010 and 011 in the previous example holds true in general.

[2] For a formal definition of graphs, see Chapter 4.



(a) $u = 010$ (b) $u = 011$

Figure 2.6: Equivalence classes of vectors $\Psi_2(u)$ and $\Psi_2(\overline{u})$ and distinct equivalence classes of vectors obtained after one or two matrix products with M_e or M_o.

Lemma 2.6.

Let u be a factor of the Thue–Morse word $\mathbf{t}$ and let $y = \Psi_2(u)$, $\overline{y} = \Psi_2(\overline{u})$. We have

$$\begin{aligned}&\{M_eM_ey, M_eM_oy, M_oM_ey, M_oM_oy, M_eM_e\overline{y}, M_eM_o\overline{y}, M_oM_e\overline{y}, M_oM_o\overline{y}\}\big/_{\sim}\\ &= \{M_eM_ey, M_eM_oy, M_oM_ey, M_oM_oy\}\big/_{\sim}\end{aligned}$$

and the cardinal of this set is equal to 2 if u starts and ends with the same letter. Otherwise, the cardinality of this set is equal to 4.

Proof. Let u be a factor of the Thue–Morse word $\mathbf{t}$ and let $y = \Psi_2(u)$, $\overline{y} = \Psi_2(\overline{u})$ with $y_i = [\Psi_2(u)]_i$ and $\overline{y}_i = [\Psi_2(\overline{u})]_i$ for $1 \leq i \leq 10$. We have

$$\begin{cases} \overline{y}_1 = y_2 \\ \overline{y}_2 = y_1 \\ \overline{y}_3 = y_6 \\ \overline{y}_4 = y_5 \\ \overline{y}_5 = y_4 \\ \overline{y}_6 = y_3 \\ \overline{y}_7 = y_{10} \\ \overline{y}_8 = y_9 \\ \overline{y}_9 = y_8 \\ \overline{y}_{10} = y_7. \end{cases}$$

Consider the set $S = \{M_eM_ey, M_eM_oy, M_oM_ey, M_oM_oy\}\big/_{\sim}$ and

$$S' = \{M_eM_ey, M_eM_oy, M_oM_ey, M_oM_oy, M_eM_e\overline{y}, M_eM_o\overline{y}, M_oM_e\overline{y}, M_oM_o\overline{y}\}\big/_{\sim}.$$

From Lemma 2.3, we have $M_eM_ey \not\sim M_oM_ey$ and $M_oM_oy \not\sim M_eM_oy$. Hence $\#S \geq 2$. Moreover, $M_eM_ey \sim M_oM_oy$ if and only if $M_eM_oy \sim M_oM_ey$. So $\#S \in \{2, 4\}$.

We have

$M_eM_ey - M_oM_oy$	$M_eM_oy - M_oM_ey$
0	0
0	0
$-y_{10} + y_2 - y_8$	$y_{10} - y_2 + y_8$
$y_1 - y_{10} + y_2 - y_7 - y_8 - y_9$	0
$y_1 - y_{10} + y_2 - y_7 - y_8 - y_9$	0
$y_1 - y_7 - y_9$	$-y_1 + y_7 + y_9$
y_7	$y_{10} + y_8$
$-y_{10}$	$-y_7 - y_9$
$-y_7$	$-y_{10} - y_8$
y_{10}	$y_7 + y_9$

and the two columns are equivalent to $0 \in \mathbb{N}^{10}$ if and only if u starts and ends with the same letter, i.e., $y_1 = y_7 + y_9$, $y_2 = y_8 + y_10$, $y_4 = y_5$.

If u starts and ends with the same letter, then applying the fourth assertion of Lemma 2.3, we have $M_ey \sim M_o\overline{y}$ and $M_oy \sim M_e\overline{y}$. So we obtain

$$M_oM_e\overline{y} \sim M_oM_oy \sim M_eM_ey \sim M_eM_o\overline{y}$$
$$\not\sim M_eM_e\overline{y} \sim M_eM_oy \sim M_oM_ey \sim M_oM_o\overline{y}.$$

In other words, the sets S and S' are equal and contain 2 elements.

Now, if u starts and ends with different letters, we have $y_1 = y_7 + y_9$, $y_2 = y_8 + y_{10}$, $y_4 - y_1 = y_5 - y_2$. Hence, $M_eM_ey \not\sim M_oM_oy$ and $M_eM_oy \not\sim M_oM_e$. Moreover, we get

$M_eM_ey - M_eM_o\overline{y}$	$M_oM_oy - M_oM_e\overline{y}$
0	0
0	0
$-y_{10} + y_2 + y_4 - y_5 - y_8$	$-y_1 + y_4 - y_5 + y_7 + y_9$
$-y_{10} + y_2 + y_4 - y_5 - y_8$	$-y_1 + y_4 - y_5 + y_7 + y_9$
$y_1 - y_4 + y_5 - y_7 - y_9$	$y_{10} - y_2 - y_4 + y_5 + y_8$
$y_1 - y_4 + y_5 - y_7 - y_9$	$y_{10} - y_2 - y_4 + y_5 + y_8$
$-y_9$	0
y_8	0
y_9	0
$-y_8$	0

and

$M_eM_oy - M_eM_e\overline{y}$	$M_oM_ey - M_oM_o\overline{y}$
0	0
0	0
$-y_1 + y_4 - y_5 + y_7 + y_9$	$-y_{10} + y_2 + y_4 - y_5 - y_8$
$-y_1 + y_4 - y_5 + y_7 + y_9$	$-y_{10} + y_2 + y_4 - y_5 - y_8$
$y_{10} - y_2 - y_4 + y_5 + y_8$	$y_1 - y_4 + y_5 - y_7 - y_9$
$y_{10} - y_2 - y_4 + y_5 + y_8$	$y_1 - y_4 + y_5 - y_7 - y_9$
y_8	0
$-y_9$	0
$-y_8$	0
y_9	0

where each column is similar to $0 \in \mathbb{N}^{10}$. Therefore, the set S is equal to S' and contains 4 elements in this case. ∎

From Lemma 2.6, the 2-abelian complexity of $\mathbf{t}$ clearly satisfies the following relation. We will see in the next subsection that it is not the only relation that holds for $\mathcal{P}_{\mathbf{t}}^{(2)}$.

Proposition 2.7.

For all $n \geq 0$, $\mathcal{P}_{\mathbf{t}}^{(2)}(2n+1) = \mathcal{P}_{\mathbf{t}}^{(2)}(4n+1)$.

2.1.2 We conjecture relations for the 2-abelian complexity using a predictive algorithm

Given the first few terms of a sequence $s(n)_{n\geq 0}$, one can easily conjecture the potential k-regularity of this sequence by exhibiting relations that should be satisfied. Allouche and Shallit gave such a "predictive" algorithm that recognizes k-regularity [AS03b, Section 6]. Their idea is to construct a list in which the elements are truncated versions of elements of the k-kernel and such that the list contains the hypothetical generators of the k-kernel. In practice, the truncated versions may contain $N = 100$ terms for a start.

To show that the 2-abelian complexity of the Thue–Morse word is 2-regular, we follow Allouche and Shallit's ideas. We compute the first $N = 100$ terms of the 63 first sequences of the 2-kernel $\mathcal{K}_2$ of $\mathcal{P}_{\mathbf{t}}^{(2)}(n)_{n\geq 0}$:

$$\begin{aligned}\mathcal{K}_2 &= \{\mathcal{P}_{\mathbf{t}}^{(2)}(2^e n + r)_{n\geq 0} : e \geq 0, 0 \leq r < 2^e\} \\ &= \{\mathcal{P}_{\mathbf{t}}^{(2)}(n)_{n\geq 0}, \mathcal{P}_{\mathbf{t}}^{(2)}(2n)_{n\geq 0}, \mathcal{P}_{\mathbf{t}}^{(2)}(2n+1)_{n\geq 0}, \mathcal{P}_{\mathbf{t}}^{(2)}(4n)_{n\geq 0} \ldots\}.\end{aligned}$$

We let $\mathbf{x}_{2^e+r}$ denote the sequence $\mathcal{P}_{\mathbf{t}}^{(2)}(2^e n + r)_{n\geq 0}$. Hence, we compute the first 100 terms of $\mathbf{x}_1, \ldots, \mathbf{x}_{63}$. In particular, to get $\mathbf{x}_{63}(100) = \mathcal{P}_{\mathbf{t}}^{(2)}(32 \cdot 99 + 31)$, we need to compute up to 3200 elements of $\mathcal{P}_{\mathbf{t}}^{(2)}$.

- At the first step $j = 1$, we select the first sequence $\mathbf{x}_1 = \mathcal{P}_{\mathbf{t}}^{(2)}(n)_{n\geq 0}$.
- At step j, there are $r < j$ previously selected sequences. We consider the sequence $\mathbf{x}_j$ and we check whether $\mathbf{x}_j$ is a linear combination of the selected sequences. If $\mathbf{x}_j$ and the selected sequences are linearly independent, then we add $\mathbf{x}_j$ to the selection.
- $j \to j + 1$ until $j < 63$.

Of course, in such an algorithm, a finite examination does not lead to a proof of the k-regularity of a sequence. We still need to verify the relations that we found through induction or other means.

Implementing this with `Mathematica`[3], we conjecture that the sequences $\mathbf{x}_{32}, \ldots, \mathbf{x}_{63}$ are all linear combinations of the sequences $\mathbf{x}_1, \ldots, \mathbf{x}_{31}$. In particular, we conjecture the

[3]The corresponding `Mathematica` code is available in Appendix A.

following relations and check them for the first 10000 terms:

$$
\begin{aligned}
\mathbf{x}_{32} &= \mathbf{x}_8\\
\mathbf{x}_{33} &= \mathbf{x}_3\\
\mathbf{x}_{34} &= \mathbf{x}_{10}\\
\mathbf{x}_{35} &= \mathbf{x}_{11}\\
\mathbf{x}_{36} &= -\mathbf{x}_{10} + \mathbf{x}_{11} + \mathbf{x}_{19}\\
\mathbf{x}_{37} &= \mathbf{x}_{19}\\
\mathbf{x}_{38} &= -\mathbf{x}_3 + \mathbf{x}_{10} + \mathbf{x}_{19}\\
\mathbf{x}_{39} &= -\mathbf{x}_3 + \mathbf{x}_{11} + \mathbf{x}_{19}\\
\mathbf{x}_{40} &= -\mathbf{x}_3 + \mathbf{x}_{10} + \mathbf{x}_{11}\\
\mathbf{x}_{41} &= \mathbf{x}_{11}\\
\mathbf{x}_{42} &= -\mathbf{x}_3 + \mathbf{x}_{10} + \mathbf{x}_{11}\\
\mathbf{x}_{43} &= -2\mathbf{x}_3 + 3\mathbf{x}_{10}\\
\mathbf{x}_{44} &= -2\mathbf{x}_3 - \mathbf{x}_6 + \mathbf{x}_7 + 3\mathbf{x}_{10}\\
\mathbf{x}_{45} &= -\mathbf{x}_3 - 3\mathbf{x}_6 + 2\mathbf{x}_7 + 3\mathbf{x}_{10} + \mathbf{x}_{11} - \mathbf{x}_{19}\\
\mathbf{x}_{46} &= -2\mathbf{x}_3 - 3\mathbf{x}_6 + 2\mathbf{x}_7 + 5\mathbf{x}_{10} + \mathbf{x}_{11} - 2\mathbf{x}_{19}\\
\mathbf{x}_{47} &= -2\mathbf{x}_3 + \mathbf{x}_7 + 3\mathbf{x}_{10} - \mathbf{x}_{19}\\
\mathbf{x}_{48} &= -\mathbf{x}_3 + \mathbf{x}_7 + \mathbf{x}_{10}\\
\mathbf{x}_{49} &= \mathbf{x}_7\\
\mathbf{x}_{50} &= -\mathbf{x}_3 + \mathbf{x}_7 + \mathbf{x}_{10}\\
\mathbf{x}_{51} &= -\mathbf{x}_3 - 3\mathbf{x}_6 + 2\mathbf{x}_7 + 3\mathbf{x}_{10} + \mathbf{x}_{11} - \mathbf{x}_{19}\\
\mathbf{x}_{52} &= -2\mathbf{x}_3 - 3\mathbf{x}_6 + 2\mathbf{x}_7 + 5\mathbf{x}_{10} + \mathbf{x}_{11} - 2\mathbf{x}_{19}\\
\mathbf{x}_{53} &= -2\mathbf{x}_3 + \mathbf{x}_7 + 3\mathbf{x}_{10} - \mathbf{x}_{19}\\
\mathbf{x}_{54} &= -4\mathbf{x}_3 + 3\mathbf{x}_6 + \mathbf{x}_7 + 3\mathbf{x}_{10} - \mathbf{x}_{11} - 2\mathbf{x}_{14} + \mathbf{x}_{15}\\
\mathbf{x}_{55} &= -4\mathbf{x}_3 + 3\mathbf{x}_6 + \mathbf{x}_7 + 3\mathbf{x}_{10} - \mathbf{x}_{11} - 3\mathbf{x}_{14} + 2\mathbf{x}_{15}\\
\mathbf{x}_{56} &= -\mathbf{x}_3 + \mathbf{x}_{10} + \mathbf{x}_{15}\\
\mathbf{x}_{57} &= \mathbf{x}_{15}\\
\mathbf{x}_{58} &= -\mathbf{x}_3 + \mathbf{x}_{10} + \mathbf{x}_{15}\\
\mathbf{x}_{59} &= -2\mathbf{x}_3 + 3\mathbf{x}_6 - \mathbf{x}_7 - \mathbf{x}_{11} + \mathbf{x}_{15} + \mathbf{x}_{19}\\
\mathbf{x}_{60} &= -4\mathbf{x}_3 + 6\mathbf{x}_6 + \mathbf{x}_{10} - 2\mathbf{x}_{11} - 3\mathbf{x}_{14} + 2\mathbf{x}_{15} + \mathbf{x}_{19}\\
\mathbf{x}_{61} &= -3\mathbf{x}_3 + 6\mathbf{x}_6 - 2\mathbf{x}_{11} - 3\mathbf{x}_{14} + 2\mathbf{x}_{15} + \mathbf{x}_{19}\\
\mathbf{x}_{62} &= -\mathbf{x}_3 + 3\mathbf{x}_6 - \mathbf{x}_7 - \mathbf{x}_{10} - \mathbf{x}_{11} + \mathbf{x}_{15} + \mathbf{x}_{19}\\
\mathbf{x}_{63} &= \mathbf{x}_{15}.
\end{aligned}
$$

Moreover, we see that not all sequences $\mathbf{x}_1, \ldots, \mathbf{x}_{31}$ occur in the previous relations. We can restrict the generators of the 2-kernel according to these conjectured relations on $\mathbf{x}_1, \ldots, \mathbf{x}_{31}$:

$$
\begin{aligned}
\mathbf{x}_5 &= \mathbf{x}_3\\
\mathbf{x}_9 &= \mathbf{x}_3\\
\mathbf{x}_{12} &= -\mathbf{x}_6 + \mathbf{x}_7 + \mathbf{x}_{11}\\
\mathbf{x}_{13} &= \mathbf{x}_7\\
\mathbf{x}_{16} &= \mathbf{x}_8\\
\mathbf{x}_{17} &= \mathbf{x}_3\\
\mathbf{x}_{18} &= \mathbf{x}_{10}\\
\mathbf{x}_{20} &= -\mathbf{x}_{10} + \mathbf{x}_{11} + \mathbf{x}_{19}\\
\mathbf{x}_{21} &= \mathbf{x}_{11}
\end{aligned}
\qquad
\begin{aligned}
\mathbf{x}_{22} &= -\mathbf{x}_3 - 2\mathbf{x}_6 + \mathbf{x}_7 + 3\mathbf{x}_{10} + \mathbf{x}_{11} - \mathbf{x}_{19}\\
\mathbf{x}_{23} &= -\mathbf{x}_3 - 3\mathbf{x}_6 + 2\mathbf{x}_7 + 3\mathbf{x}_{10} + \mathbf{x}_{11} - \mathbf{x}_{19}\\
\mathbf{x}_{24} &= -\mathbf{x}_3 + \mathbf{x}_7 + \mathbf{x}_{10}\\
\mathbf{x}_{25} &= \mathbf{x}_7\\
\mathbf{x}_{26} &= -\mathbf{x}_3 + \mathbf{x}_7 + \mathbf{x}_{10}\\
\mathbf{x}_{27} &= -2\mathbf{x}_3 + \mathbf{x}_7 + 3\mathbf{x}_{10} - \mathbf{x}_{19}\\
\mathbf{x}_{28} &= -2\mathbf{x}_3 + \mathbf{x}_7 + 3\mathbf{x}_{10} - \mathbf{x}_{14} + \mathbf{x}_{15} - \mathbf{x}_{19}\\
\mathbf{x}_{29} &= \mathbf{x}_{15}\\
\mathbf{x}_{30} &= -\mathbf{x}_3 + 3\mathbf{x}_6 - \mathbf{x}_7 - \mathbf{x}_{10} - \mathbf{x}_{11} + \mathbf{x}_{15} + \mathbf{x}_{19}\\
\mathbf{x}_{31} &= -3\mathbf{x}_3 + 6\mathbf{x}_6 - 2\mathbf{x}_{11} - 3\mathbf{x}_{14} + 2\mathbf{x}_{15} + \mathbf{x}_{19}.
\end{aligned}
$$

If the conjectured relations hold, then any sequence $\mathbf{x}_n$ for $n \geq 32$ is a linear combination of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{19}$. Indeed, consider a sequence $\mathbf{x}_{2^e+r} = \mathcal{P}_{\mathbf{t}}^{(2)}(2^e n + r)_{n\geq 0}$ with $e \geq 5$. Then $\mathcal{P}_{\mathbf{t}}^{(2)}(2^e n+r)$ can be written as $\mathcal{P}_{\mathbf{t}}^{(2)}(32n'+r')$ with $0 \leq r' < 32$ and we can apply the relation corresponding to $\mathbf{x}_{32+r'}$.

Example 2.8. Consider the sequence $\mathbf{x}_{154} = \mathcal{P}_{\mathbf{t}}^{(2)}(128n + 26)_{n\geq 0}$ of the 2-kernel. We have

$$
\begin{aligned}
\mathcal{P}_{\mathbf{t}}^{(2)}(128n + 26) = \mathcal{P}_{\mathbf{t}}^{(2)}(32(4n) + 26)&\\
&= -\mathcal{P}_{\mathbf{t}}^{(2)}(2(4n) + 1) + \mathcal{P}_{\mathbf{t}}^{(2)}(8(4n) + 2) + \mathcal{P}_{\mathbf{t}}^{(2)}(8(4n) + 7)\\
&= -\mathcal{P}_{\mathbf{t}}^{(2)}(8n + 1) + \mathcal{P}_{\mathbf{t}}^{(2)}(32n + 2) + \mathcal{P}_{\mathbf{t}}^{(2)}(32n + 7)
\end{aligned}
$$

as $\mathcal{P}_{\mathbf{t}}^{(2)}(32n + 26)_{n\geq 0} = \mathbf{x}_{58} = -\mathbf{x}_3 + \mathbf{x}_{10} + \mathbf{x}_{15}$. So we get

$$
\begin{aligned}
\mathbf{x}_{154} &= -\mathbf{x}_9 + \mathbf{x}_{34} + \mathbf{x}_{39}\\
&= -\mathbf{x}_3 + \mathbf{x}_{10} - \mathbf{x}_3 + \mathbf{x}_{11} + \mathbf{x}_{19} &&\text{(using the conjectured relations)}\\
&= -2\mathbf{x}_3 + \mathbf{x}_{10} + \mathbf{x}_{11} + \mathbf{x}_{19}.
\end{aligned}
$$

Note that this is not the only way to obtain a linear combination of $\mathbf{x}_1, \ldots, \mathbf{x}_{19}$. For example, we could choose to write $\mathcal{P}^{(2)}(128n+26) = \mathcal{P}^{(2)}(16(8n+1)+10)$ and then proceed as before.

Let us consider another example with the sequence $\mathbf{x}_{164} = \mathcal{P}_{\mathbf{t}}^{(2)}(128n + 36)_{n\geq 0}$. We have that $\mathcal{P}_{\mathbf{t}}^{(2)}(128n + 36) = \mathcal{P}_{\mathbf{t}}^{(2)}(32(4n + 1) + 4)$. Then, using the conjectured relation $\mathbf{x}_{36} = -\mathbf{x}_{10} + \mathbf{x}_{11} + \mathbf{x}_{19}$, we obtain that $\mathcal{P}_{\mathbf{t}}^{(2)}(128n + 36)$ is equal to

$$
\begin{aligned}
-\mathcal{P}_{\mathbf{t}}^{(2)}&(8(4n + 1) + 2) + \mathcal{P}_{\mathbf{t}}^{(2)}(8(4n + 1) + 3) + \mathcal{P}_{\mathbf{t}}^{(2)}(16(4n + 1) + 3)\\
&= -\mathcal{P}_{\mathbf{t}}^{(2)}(32n + 10) + \mathcal{P}_{\mathbf{t}}^{(2)}(32n + 11) + \mathcal{P}_{\mathbf{t}}^{(2)}(32(2n) + 19)\\
&= -\mathcal{P}_{\mathbf{t}}^{(2)}(32n + 10) + \mathcal{P}_{\mathbf{t}}^{(2)}(32n + 11) - \mathcal{P}_{\mathbf{t}}^{(2)}(2(2n) + 1) + 3\mathcal{P}_{\mathbf{t}}^{(2)}(4(2n) + 2)\\
&\quad + 2\mathcal{P}_{\mathbf{t}}^{(2)}(4(2n) + 3) - 3\mathcal{P}_{\mathbf{t}}^{(2)}(8(2n) + 2) + \mathcal{P}_{\mathbf{t}}^{(2)}(8(2n) + 3) - \mathcal{P}_{\mathbf{t}}^{(2)}(16(2n) + 3)\\
&= -\mathcal{P}_{\mathbf{t}}^{(2)}(32n + 10) + \mathcal{P}_{\mathbf{t}}^{(2)}(32n + 11) - \mathcal{P}_{\mathbf{t}}^{(2)}(4n + 1) + 3\mathcal{P}_{\mathbf{t}}^{(2)}(8n + 2)\\
&\quad + 2\mathcal{P}_{\mathbf{t}}^{(2)}(8n + 3) - 3\mathcal{P}_{\mathbf{t}}^{(2)}(16n + 2) + \mathcal{P}_{\mathbf{t}}^{(2)}(16n + 3) - \mathcal{P}_{\mathbf{t}}^{(2)}(32n + 3)
\end{aligned}
$$

where we used the conjectured relation $\mathbf{x}_{51} = -\mathbf{x}_3 + 3\mathbf{x}_6 + 2\mathbf{x}_7 + 3\mathbf{x}_{10} + \mathbf{x}_{11} - \mathbf{x}_{19}$ to obtain the third equality.

Finally, we get

$$\begin{aligned}
\mathbf{x}_{164} &= -\mathbf{x}_{42} + \mathbf{x}_{43} - \mathbf{x}_5 + 3\mathbf{x}_{10} + 2\mathbf{x}_{11} - 3\mathbf{x}_{18} + \mathbf{x}_{19} - \mathbf{x}_{35} \\
&= -(-\mathbf{x}_3 + \mathbf{x}_{10} + \mathbf{x}_{11}) + (-2\mathbf{x}_3 + 3\mathbf{x}_{10}) - \mathbf{x}_3 + 3\mathbf{x}_{10} + 2\mathbf{x}_{11} - 3\mathbf{x}_{10} + \mathbf{x}_{19} - \mathbf{x}_{11} \\
&= -2\mathbf{x}_3 + 2\mathbf{x}_{10} + \mathbf{x}_{19}.
\end{aligned}$$

In other words, if the conjectured relations hold, the 2-kernel of the 2-abelian complexity of the Thue–Morse word is finitely generated by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{19}$. As a direct consequence of Proposition 2.7, some of the conjectured relations are proved.

Corollary 2.9.

Let $\mathbf{x}_{2^e+r}$ denote the subsequence $\mathcal{P}_{\mathbf{t}}^{(2)}(2^e n + r)_{n\geq 0}$ of the 2-abelian complexity of the Thue–Morse word, with $e \geq 0, 0 \leq r < 2^e$. We have

$$\begin{aligned}
\mathbf{x}_3 &= \mathbf{x}_5 = \mathbf{x}_9 = \mathbf{x}_{17} = \mathbf{x}_{33} & \mathbf{x}_{19} &= \mathbf{x}_{37} \\
\mathbf{x}_7 &= \mathbf{x}_{13} = \mathbf{x}_{25} = \mathbf{x}_{49} & \mathbf{x}_{23} &= \mathbf{x}_{45} \\
\mathbf{x}_{11} &= \mathbf{x}_{21} = \mathbf{x}_{41} & \mathbf{x}_{27} &= \mathbf{x}_{53} \\
\mathbf{x}_{15} &= \mathbf{x}_{29} = \mathbf{x}_{57} & \mathbf{x}_{31} &= \mathbf{x}_{61}.
\end{aligned}$$

Recently, Greinecker proved our conjectured relations and so that the 2-abelian complexity of $\mathbf{t}$ is 2-regular [Gre]. His proof is based on reading frames, which are particular partitions of words, and a variation of reading frames, called off-beat frames. In particular, he reduces the study of $\mathcal{P}_{\mathbf{t}}^{(2)}$ to the study of some vectors. Instead of considering vectors of length 10 as we did with $\Psi_2(w)$, he considers vectors $\mathrm{vec}(w)$ of length 3:

$$\mathrm{vec} : \mathrm{Fac}(\mathbf{t}) \to \{\mathtt{0}, \mathtt{1}\} \times \mathbb{N} \times \{0, 1\}, w \mapsto \begin{pmatrix} \mathrm{pref}_1(w) \\ p(w) \\ r(w) \end{pmatrix}$$

where $p(w) = |w|_{\mathtt{00}} + |w|_{\mathtt{11}}$ counts the number of pairs occurring in w and $r(w)$ determines whether $\mathrm{pref}_2(w)$ is in the off-beat frame or not.

Theorem 2.10. [Gre]

Let $\mathbf{x}_{2^e+r}$ denote the subsequence $\mathcal{P}_{\mathbf{t}}^{(2)}(2^e n + r)_{n\geq 0}$ of the 2-abelian complexity of the Thue–Morse word, with $e \geq 0, 0 \leq r < 2^e$. We have

$$\begin{aligned}
\mathbf{x}_5 &= \mathbf{x}_3\\
\mathbf{x}_{12} &= -\mathbf{x}_6 + \mathbf{x}_7 + \mathbf{x}_{11}\\
\mathbf{x}_{16} &= \mathbf{x}_8\\
\mathbf{x}_{18} &= \mathbf{x}_{10}\\
\mathbf{x}_{22} &= -\mathbf{x}_3 - 2\mathbf{x}_6 + \mathbf{x}_7 + 3\mathbf{x}_{10} + \mathbf{x}_{11} - \mathbf{x}_{19}\\
\mathbf{x}_{23} &= -\mathbf{x}_3 - 3\mathbf{x}_6 + 2\mathbf{x}_7 + 3\mathbf{x}_{10} + \mathbf{x}_{11} - \mathbf{x}_{19}\\
\mathbf{x}_{24} &= -\mathbf{x}_3 + \mathbf{x}_7 + \mathbf{x}_{10}\\
\mathbf{x}_{26} &= -\mathbf{x}_3 + \mathbf{x}_7 + \mathbf{x}_{10}\\
\mathbf{x}_{27} &= -2\mathbf{x}_3 + \mathbf{x}_7 + 3\mathbf{x}_{10} - \mathbf{x}_{19}\\
\mathbf{x}_{30} &= -\mathbf{x}_3 + 3\mathbf{x}_6 - \mathbf{x}_7 - \mathbf{x}_{10} - \mathbf{x}_{11} + \mathbf{x}_{15} + \mathbf{x}_{19}\\
\mathbf{x}_{31} &= -\mathbf{x}_3 + 6\mathbf{x}_6 - 2\mathbf{x}_{11} - 3\mathbf{x}_{14} + 2\mathbf{x}_{15} + \mathbf{x}_{19}\\
\mathbf{x}_{35} &= \mathbf{x}_{11}\\
\mathbf{x}_{51} &= -\mathbf{x}_3 - 3\mathbf{x}_6 + 2\mathbf{x}_7 + 3\mathbf{x}_{10} + \mathbf{x}_{11} - \mathbf{x}_{19}.
\end{aligned}$$

Moreover the 2-abelian complexity of the Thue–Morse word is 2-regular.

Along the way, he shows that the 2-abelian complexity of $\mathbf{t}$ is a concatenation of longer and longer palindromes:

$$\begin{aligned}
\mathcal{P}_{\mathbf{t}}^{(2)}(n)_{n\geq 0} = (1,2,4,6,8,\,&\underline{6,8,10,8,6},\underline{8,8,10,10,10,8,8},\\
&\underline{6,8,10,10,8,10,12,12,10,12,12,10,8,10,10,8,6},\ldots).
\end{aligned}$$

Otherwise stated, the graph of the values of $\mathcal{P}_{\mathbf{t}}^{(2)}$ satisfies a reflection symmetry. We obtain the same result with our method in Proposition 2.37.

2.2 Sequences satisfying a reflection symmetry

If we sketch the first values of 2-abelian complexity of the Thue–Morse word (Figures 2.7 and 2.8), we observe that the sequence seems to satisfy a reflection symmetry in the values taken over intervals of the form $[2^\ell + 1, 2^{\ell+1} + 1]$ with $\ell \geq 1$.

Other sequences satisfy a reflection symmetry over intervals of the form $[2^\ell, 2^{\ell+1}]$. For instance, it is the case of the abelian complexity of the 2-block coding of the period-doubling word $\mathbf{p}$ introduced in Example 1.16. (The recurrence satisfied by this sequence is given in Theorem 2.42.) Some values of this sequence are depicted in Figures 2.9 and 2.10.

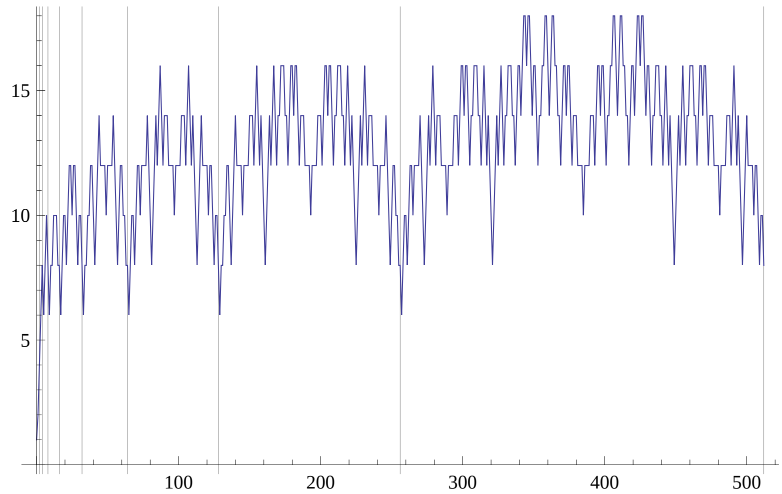


Figure 2.7: The 2-abelian complexity of $\mathbf{t}$ on the interval $[0, 512]$.

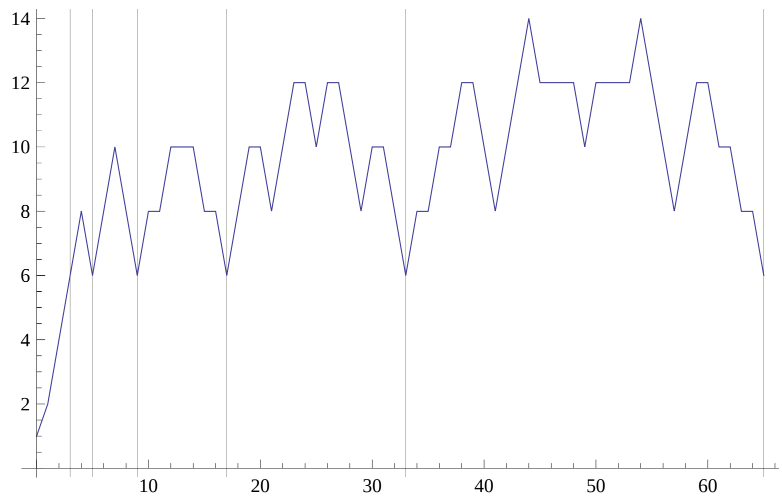


Figure 2.8: The 2-abelian complexity of $\mathbf{t}$ on the interval $[0, 64]$.

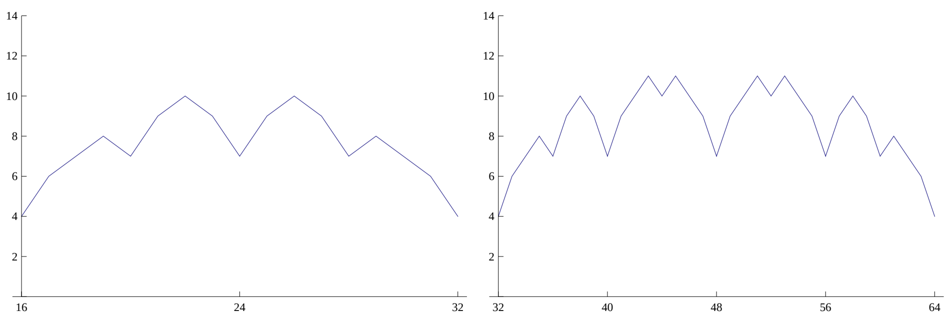


Figure 2.9: The abelian complexity of block$(\mathbf{p}, 2)$ on the intervals $[16, 32]$ and $[32, 64]$.

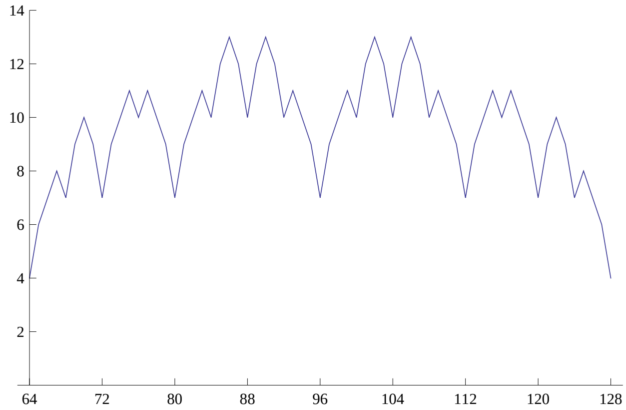


Figure 2.10: The abelian complexity of block$(\mathbf{p}, 2)$ on the interval $[64, 128]$.

In this framework, we are able to prove a general regularity result. Sequences satisfying a recurrence relation of the following form are 2-regular.

Theorem 2.11.

Let $\ell_0 \geq 0$ and $c \in \mathbb{Z}$. Suppose $s(n)_{n\geq 0}$ is a sequence such that, for all $\ell \geq \ell_0$ and $0 \leq r \leq 2^\ell - 1$, we have

$$s(2^\ell + r) = \begin{cases} s(r) + c & \text{if } r \leq 2^{\ell-1} \\ s(2^{\ell+1} - r) & \text{if } r > 2^{\ell-1}. \end{cases} \tag{2.1}$$

Then $s(n)_{n\geq 0}$ is 2-regular.

The recurrence satisfied by $s(n)$ in Theorem 2.11 reads words from left to right, i.e., starting with the most significant digit. Our proof (given page 67) of this theorem will express sequences in the 2-kernel of $s(n)_{n\geq 0}$ as in Definition 1.26, starting with the least significant digit.

Remark 2.12. From Equation (2.1) one can get some information about the asymptotic behaviour of the sequence $s(n)_{n\geq 0}$. We have $s(n) = O(\log n)$, and moreover

$$s\left(\tfrac{4^{\ell+1}-1}{3}\right) = s(4^\ell + \cdots + 4^1 + 4^0) = \left(\ell - \left\lfloor \tfrac{\ell_0 - 1}{2} \right\rfloor\right) c + s\left(\tfrac{4^{\lfloor(\ell_0+1)/2\rfloor} - 1}{3}\right)$$

for $\ell \geq \lfloor \frac{\ell_0 - 1}{2} \rfloor$. At the same time, there are many subsequences of $s(n)_{n\geq 0}$ which are constant; for example, $s(2^\ell) = c$ for $\ell \geq \ell_0$.

Before proving Theorem 2.11 in generality, we first examine the sequence satisfying the recurrence for $\ell_0 = 0$ and $c = 1$. It will turn out that the general solution can be expressed naturally in terms of this sequence.

Let $A(0) = 0$. For each $\ell \geq 0$ and $0 \leq r \leq 2^\ell - 1$, let

$$A(2^\ell + r) = \begin{cases} A(r) + 1 & \text{if } r \leq 2^{\ell-1} \\ A(2^{\ell+1} - r) & \text{if } r > 2^{\ell-1}. \end{cases} \tag{2.2}$$

The sequence $A(n)_{n\geq 0}$ is

$$0, 1, 1, 2, 1, 2, 2, 2, 1, 2, 2, 3, 2, 3, 2, 2, \ldots$$

and appears as A007302 in [OF]. Allouche and Shallit [AS03b, Example 12] identified this sequence as an example of a regular sequence. We include a proof here.

Proposition 2.13.

For all $n \geq 0$, we have

$$\begin{aligned} A(2n) &= A(n) \\ A(8n+1) &= A(4n+1) \\ A(8n+3) &= A(2n+1)+1 \\ A(8n+5) &= A(2n+1)+1 \\ A(8n+7) &= A(4n+3). \end{aligned}$$

In particular, $A(n)_{n\geq 0}$ is 2-regular.

Proof. This proof is typical of many of the proofs throughout the chapter. We work by induction on n. The case $n = 0$ can be checked easily using the first few values of the sequence $A(n)_{n\geq 0}$. Therefore, let $n \geq 1$ and assume that the recurrence holds for all values less than n. Write $n = 2^\ell + r$ with $\ell \geq 0$ and $0 \leq r \leq 2^\ell - 1$.

First let us address the equation $A(2n) = A(n)$. If $0 \leq r \leq 2^{\ell-1}$, then

$$\begin{aligned} A(2n) &= A(2^{\ell+1} + 2r) \\ &= A(2r) + 1 && \text{(by Equation (2.2))} \\ &= A(r) + 1 && \text{(by inductive hypothesis)} \\ &= A(2^\ell + r) && \text{(by Equation (2.2))} \\ &= A(n). \end{aligned}$$

On the other hand, if $2^{\ell-1} < r < 2^\ell$, then

$$\begin{aligned} A(2n) &= A(2^{\ell+1} + 2r) \\ &= A(2^{\ell+2} - 2r) && \text{(by Equation (2.2))} \\ &= A(2^{\ell+1} - r) && \text{(by inductive hypothesis)} \\ &= A(2^\ell + r) && \text{(by Equation (2.2))} \\ &= A(n). \end{aligned}$$

Next we consider $A(8n+1) = A(4n+1)$. If $0 \leq r \leq 2^{\ell-1} - 1$, then

$$\begin{aligned} A(8n+1) &= A(2^{\ell+3} + 8r + 1) \\ &= A(8r+1) + 1 && \text{(by Equation (2.2))} \\ &= A(4r+1) + 1 && \text{(by inductive hypothesis)} \\ &= A(2^{\ell+2} + 4r + 1) && \text{(by Equation (2.2))} \\ &= A(4n+1). \end{aligned}$$

If $2^{\ell-1} \le r < 2^{\ell}$, then

$$\begin{aligned}
A(8n+1) &= A(2^{\ell+3}+8r+1) \\
&= A(2^{\ell+4}-8r-1) && \text{(by Equation (2.2))} \\
&= A(2^{\ell+4}-8r-8+7) \\
&= A(2^{\ell+3}-4r-4+3) && \text{(by inductive hypothesis)} \\
&= A(2^{\ell+3}-(4r+1)) \\
&= A(2^{\ell+2}+4r+1) && \text{(by Equation (2.2))} \\
&= A(4n+1).
\end{aligned}$$

The equations for $A(8n+3)$, $A(8n+5)$ and $A(8n+7)$ are handled similarly. □

Now we prove Theorem 2.11. We show that for general $\ell_0 \ge 0$, a sequence $s(n)_{n\ge 0}$ satisfying the recurrence can be written in terms of $A(n)_{n\ge 0}$.

Proof of Theorem 2.11. There are 2^{ℓ_0} initial conditions for the recurrence, namely $s(0), \dots, s(2^{\ell_0}-1)$. We claim that most of the 2^{ℓ_0+2} subsequences of the form $s(2^{\ell_0+2}n+i)_{n\ge 0}$ depend on only one of the initial conditions $s(j)$; each of these subsequences is essentially $A(n)_{n\ge 0}$, $A(4n+1)_{n\ge 0}$, $A(2n+1)_{n\ge 0}$, or $A(4n+3)_{n\ge 0}$. Furthermore, each of the remaining subsequences is equal to $s(2^{\ell_0}n+j)+c$ for some j. More precisely, for $0 \le i \le 2^{\ell_0+2}-1$ and $n \ge 0$ we have the identity

$$s(2^{\ell_0+2}n+i) = \\
\begin{cases}
c\,A(n)+s(0) & \text{if } i=0 \\
c\,A(4n+1)-c+s(i) & \text{if } 1 \le i \le 2^{\ell_0}-1 \\
c\,A(4n+1)+s(0) & \text{if } i=2^{\ell_0} \\
s(2^{\ell_0}n+i-2^{\ell_0})+c & \text{if } 2^{\ell_0}+1 \le i \le 2^{\ell_0}+2^{\ell_0-1}-1 \\
c\,A(2n+1)+s(|i-2^{\ell_0+1}|) & \text{if } 2^{\ell_0}+2^{\ell_0-1} \le i \le 2^{\ell_0+1}+2^{\ell_0-1} \\
s(2^{\ell_0}n+i-2^{\ell_0+1})+c & \text{if } 2^{\ell_0+1}+2^{\ell_0-1}+1 \le i \le 2^{\ell_0+1}+2^{\ell_0}-1 \\
c\,A(4n+3)+s(0) & \text{if } i=2^{\ell_0+1}+2^{\ell_0} \\
c\,A(4n+3)-c+s(2^{\ell_0+2}-i) & \text{if } 2^{\ell_0+1}+2^{\ell_0}+1 \le i \le 2^{\ell_0+2}-1.
\end{cases}$$

(Note the symmetry among the eight cases, which reflects the symmetry $s(2^{\ell}+r) = s(2^{\ell+1}-r)$ of the recurrence for $r > 2^{\ell-1}$.) It will follow from this identity that the $\mathbb{Z}$-module generated by the 2-kernel of $s(n)_{n\ge 0}$ is generated by the sequences $s(2^{\ell}n+j)_{n\ge 0}$ for $0 \le \ell \le \ell_0+1$ and $0 \le j \le 2^{\ell}-1$, $A(n)_{n\ge 0}$, $A(4n+1)_{n\ge 0}$, $A(2n+1)_{n\ge 0}$, $A(4n+3)_{n\ge 0}$, and the constant 1 sequence. In particular, this module is finitely generated.

We prove the identity by induction on n. Recall that for all $\ell \ge \ell_0$ and $0 \le r \le 2^{\ell}-1$ we have Equation (2.1), i.e.,

$$s(2^{\ell}+r) = \begin{cases} s(r)+c & \text{if } r \le 2^{\ell-1} \\ s(2^{\ell+1}-r) & \text{if } r > 2^{\ell-1}. \end{cases}$$

For $n=0$, one uses $A(1)=1$ and $A(3)=2$ to verify that all eight cases of the identity hold. Inductively, let $n \ge 1$, and assume the identity is true for all $n' < n$. Write $n = 2^{\ell}+r$ with $\ell \ge 0$ and $0 \le r \le 2^{\ell}-1$.

First we consider the case $0 \le r \le 2^{\ell-1} - 1$. For all $0 \le i \le 2^{\ell_0+2} - 1$ we have $2^{\ell_0+2} r + i \le 2^{(\ell_0+2+\ell)-1} - 1$, so

$$\begin{aligned} s(2^{\ell_0+2} n + i) &= s(2^{\ell_0+2+\ell} + (2^{\ell_0+2} r + i)) \\ &= s(2^{\ell_0+2} r + i) + c \qquad \text{(by Equation (2.1)).} \end{aligned}$$

If $1 \le i \le 2^{\ell_0} - 1$, then the inductive hypothesis now gives

$$\begin{aligned} s(2^{\ell_0+2} n + i) &= s(2^{\ell_0+2} r + i) + c \\ &= c\, A(4r + 1) + s(i) \\ &= c\, \big(A(2^{\ell+2} + 4r + 1) - 1\big) + s(i) \\ &= c\, A(4n + 1) - c + s(i), \end{aligned}$$

where we have used $A(2^{\ell+2} + 4r + 1) = A(4r + 1) + 1$ from the recurrence for $A(n)$, since $4r + 1 \le 2^{(\ell+2)-1}$. The other seven intervals for i are verified similarly; in each case one applies the inductive hypothesis to $s(2^{\ell_0+2} r + i) + c$ and then uses the recurrence for either $A(n)$ or $s(n)$ to raise an argument in r to an argument in n.

It remains to consider $2^{\ell-1} \le r \le 2^{\ell} - 1$. First we address the case $i = 0$. If $r = 2^{\ell-1}$, then

$$\begin{aligned} s(2^{\ell_0+2} n + i) &= s(2^{\ell_0+2+\ell} + 2^{\ell_0+2+\ell-1}) \\ &= s(2^{\ell_0+2+\ell-1}) + c && \text{(by Equation (2.1))} \\ &= c\, A(2^{\ell-1}) + s(0) + c && \text{(by inductive hypothesis)} \\ &= c\, \big(A(2^{\ell} + 2^{\ell-1}) - 1\big) + s(0) + c && \text{(by Equation (2.2))} \\ &= c\, A(n) + s(0) \end{aligned}$$

as desired. Alternatively, if $2^{\ell-1} < r \le 2^{\ell} - 1$ then $2^{\ell_0+2} r > 2^{(\ell_0+2+\ell)-1}$, so

$$\begin{aligned} s(2^{\ell_0+2} n + i) &= s(2^{\ell_0+2+\ell} + 2^{\ell_0+2} r) \\ &= s(2^{\ell_0+2+\ell+1} - 2^{\ell_0+2} r) && \text{(by Equation (2.1))} \\ &= s(2^{\ell_0+2}(2^{\ell+1} - r) + 0) \\ &= c\, A(2^{\ell+1} - r) + s(0) && \text{(by inductive hypothesis)} \\ &= c\, A(2^{\ell} + r) + s(0) && \text{(by Equation (2.2))} \\ &= c\, A(n) + s(0). \end{aligned}$$

Therefore it remains to consider $2^{\ell-1} \le r \le 2^{\ell} - 1$ for $1 \le i \le 2^{\ell_0+2} - 1$. In this range we have $2^{\ell_0+2} r + i > 2^{(\ell_0+2+\ell)-1}$, so

$$\begin{aligned} s(2^{\ell_0+2} n + i) &= s(2^{\ell_0+2+\ell} + (2^{\ell_0+2} r + i)) \\ &= s(2^{\ell_0+2+\ell+1} - 2^{\ell_0+2} r - i) && \text{(by Equation (2.1))} \\ &= s(2^{\ell_0+2} n' + i'), \end{aligned}$$

where $n' = 2^{\ell+1} - r - 1$ and $i' = 2^{\ell_0+2} - i$. We prove the identity for the seven intervals for i using the same steps we have already used several times; we have just applied the recurrence for $s(n)$, so next we use the inductive hypothesis, followed by the recurrence for $A(n)$ or

$s(n)$, depending on which term appears. For the first interval, if $1 \leq i \leq 2^{\ell_0} - 1$, then $2^{\ell_0+1} + 2^{\ell_0} + 1 \leq i' \leq 2^{\ell_0+2} - 1$, so

$$\begin{aligned} s(2^{\ell_0+2}n + i) &= s(2^{\ell_0+2}n' + i') \\ &= c\,A(4n' + 3) - c + s(2^{\ell_0+2} - i') && \text{(by inductive hypothesis)} \\ &= c\,A(2^{\ell+3} - (4r + 1)) - c + s(i) \\ &= c\,A(2^{\ell+2} + 4r + 1) - c + s(i) && \text{(by Equation (2.2))} \\ &= c\,A(4n + 1) - c + s(i). \end{aligned}$$

The proofs for the remaining six intervals are routine at this point, so we omit the steps here. □

Example 2.14. In Section 2.3, we will use Theorem 2.11 with $\ell_0 = 1$ to conclude that $\Delta_{12}(n)_{n\geq 0}$ is 2-regular for the Thue–Morse word. In Section 2.4, we will use Theorem 2.11 with $\ell_0 = 2$ to conclude that $\Delta_0(n)_{n\geq 0}$ and $\mathcal{P}^{(1)}_{\mathbf{x}}(n)_{n\geq 0}$ are 2-regular for the period-doubling word. For $\ell_0 = 2$ the value of $s(16n + i)$ is

$$s(16n+i) = \begin{cases} c\,A(n) + s(0) & \text{if } i = 0 \\ c\,A(4n+1) - c + s(i) & \text{if } 1 \leq i \leq 3 \\ c\,A(4n+1) + s(0) & \text{if } i = 4 \\ s(4n+1) + c & \text{if } i = 5 \\ c\,A(2n+1) + s(|i-8|) & \text{if } 6 \leq i \leq 10 \\ s(4n+3) + c & \text{if } i = 11 \\ c\,A(4n+3) + s(0) & \text{if } i = 12 \\ c\,A(4n+3) - c + s(16-i) & \text{if } 13 \leq i \leq 15. \end{cases}$$

2.3 The case of the Thue–Morse word t

As noticed in the previous section, the graph of the 2-abelian complexity of the Thue–Morse word $\mathbf{t}$ introduced in Example 1.50 exhibits a reflection symmetry, but we will see that it does not satisfy recurrence relations of the type given in Theorem 2.11.

First we consider the 2-block coding of $\mathbf{t}$ denoted by $\mathbf{y}$,

$$\mathbf{y} := \mathrm{block}(\mathbf{t}, 2) = \texttt{132120132012132120121320}\cdots$$

and its abelian complexity (Figure 2.11). Recall that $\mathbf{y}$ is a fixed point of the morphism ν defined by $\nu : \texttt{0} \mapsto \texttt{12}, \texttt{1} \mapsto \texttt{13}, \texttt{2} \mapsto \texttt{20}, \texttt{3} \mapsto \texttt{21}$. It will turn out that the abelian complexity of $\mathbf{y}$ satisfies recurrence relations that are similar to Theorem 2.11 (given in Theorem2.28) and is actually 2-regular.

In order to show that the abelian complexity of the ordinary paperfolding word is 2-regular, Madill and Rampersad made a substantial use of a function computing the difference between the number of 0's and the number of 1's in factors of given length [MR13]. We will follow their lead and make use of functions related to the number of 1's and 2's (or, equivalently, the total number of 0's and 3's) in the factors of $\mathbf{y}$ of a fixed length. We will

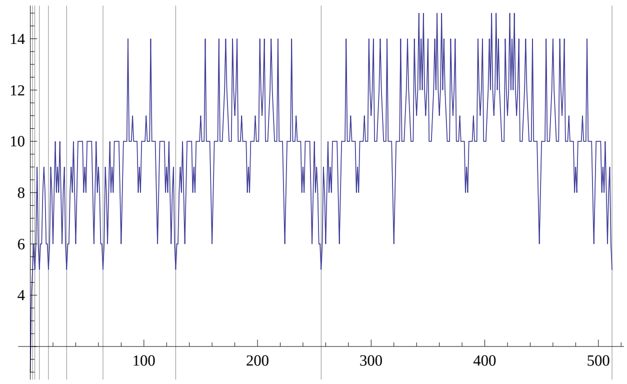


Figure 2.11: The abelian complexity of the 2-block coding $\mathbf{y}$ of the Thue–Morse word $\mathbf{t}$.

show in Lemma 2.18 that the letters $\mathtt{1}$ and $\mathtt{2}$ alternate in $\mathbf{y}$. Therefore, for $n \in \mathbb{N}$, we set

$$\begin{aligned} \max\nolimits_{12}(n) &:= \max\{|u|_{\mathtt{1}} + |u|_{\mathtt{2}} : u \text{ is a factor of } \mathbf{y} \text{ with } |u| = n\}, \\ \min\nolimits_{12}(n) &:= \min\{|u|_{\mathtt{1}} + |u|_{\mathtt{2}} : u \text{ is a factor of } \mathbf{y} \text{ with } |u| = n\}, \\ \Delta_{12}(n) &:= \max\nolimits_{12}(n) - \min\nolimits_{12}(n). \end{aligned}$$

Remark 2.15. Each of the $\Delta_{12}(n)+1$ integers in the interval $[\min_{12}(n), \max_{12}(n)]$ is attained as the number of $\mathtt{1}$'s and $\mathtt{2}$'s in some factor of $\mathbf{y}$ of length n. Indeed, when we slide a window of length n along $\mathbf{y}$ from a factor with $\min_{12}(n)$ ones and twos to a factor with $\max_{12}(n)$ ones and twos, the number of $\mathtt{1}$'s and $\mathtt{2}$'s changes by at most 1 per step.

Firstly, we prove that the abelian complexity of $\mathbf{y}$ is 2-regular by proving it is piecewise-defined together with the fact that $\Delta_{12}(n)_{n\geq 0}$ is 2-regular and the predicates occurring in the expression of $\mathcal{P}^{(1)}_{\mathbf{y}}$ are 2-automatic. In particular, we show that $\Delta_{12}(n)_{n\geq 0}$ satisfies a reflection symmetry. This permits us to express recurrence relations for $\mathcal{P}^{(1)}_{\mathbf{y}}$ at the end of Subsection 2.3.1. Secondly, we compare $\mathcal{P}^{(2)}_{\mathbf{t}}(n+1)$ and $\mathcal{P}^{(1)}_{\mathbf{y}}(n)$. We then show that the 2-regularity of $\Delta_{12}(n)_{n\geq 0}$ and $\mathcal{P}^{(1)}_{\mathbf{y}}(n)_{n\geq 0}$ implies the 2-regularity of $\mathcal{P}^{(2)}_{\mathbf{t}}(n)_{n\geq 0}$.

2.3.1 The abelian complexity of $\mathrm{block}(\mathbf{t}, 2)$ is 2-regular

The fact that $\mathcal{P}^{(1)}_{\mathbf{y}}(n)_{n\geq 0}$ is 2-regular will follow from the next statement.

Proposition 2.16.

Let $n \in \mathbb{N}$. We have

$$\mathcal{P}^{(1)}_{\mathbf{y}}(n) = \begin{cases} 2\Delta_{12}(n) + 2 & \text{if } n \text{ is odd} \\ \frac{5}{2}\Delta_{12}(n) + \frac{5}{2} & \text{if } n \text{ and } \Delta_{12}(n) + 1 \text{ are even} \\ \frac{5}{2}\Delta_{12}(n) + 4 & \text{if } n,\ \Delta_{12}(n) \text{ and } \min_{12}(n) + 1 \text{ are even} \\ \frac{5}{2}\Delta_{12}(n) + 1 & \text{if } n,\ \Delta_{12}(n) \text{ and } \min_{12}(n) \text{ are even.} \end{cases}$$

To be able to apply the composition result given by Lemma 1.34 to the expression of $\mathcal{P}_{\mathbf{y}}^{(1)}$ derived in Proposition 2.16, we have therefore to prove that

- the sequence $\Delta_{12}(n)_{n\geq 0}$ is 2-regular and
- the predicates occurring in the expression of $\mathcal{P}_{\mathbf{y}}^{(1)}$ are 2-automatic.

We first need three technical lemmas about factors of $\mathbf{y} = \mathrm{block}(\mathbf{t}, 2)$.

Lemma 2.17.

The set $\mathrm{Fac}_2(\mathbf{y})$ of factors of length 2 occurring in $\mathbf{y}$ is $\{\mathtt{01}, \mathtt{12}, \mathtt{13}, \mathtt{20}, \mathtt{21}, \mathtt{32}\}$.

Proof. It is easy to check that these six words are factors:

$$\mathrm{block}(\mathbf{t}, 2) = \underline{\mathtt{13}}\mathtt{212}\underline{\mathtt{01}}\mathtt{320}\underline{\mathtt{12}}\mathtt{1}\underline{\mathtt{32}}\mathtt{1201}\underline{\mathtt{21}}\mathtt{3}\underline{\mathtt{20}}\cdots$$

To prove that they are the only ones, it is enough to check that for any element u in $\{\mathtt{01}, \mathtt{12}, \mathtt{13}, \mathtt{20}, \mathtt{21}, \mathtt{32}\}$, the three factors of length 2 of $\nu(u)$ are still in $\{\mathtt{01}, \mathtt{12}, \mathtt{13}, \mathtt{20}, \mathtt{21}, \mathtt{32}\}$. □

The following lemma has already been observed in [KSZ, Lemma 10].

Lemma 2.18.

If w is a factor of $\mathbf{y}$, then $\big||w|_{\mathtt{1}} - |w|_{\mathtt{2}}\big| \leq 1$ and $\big||w|_{\mathtt{0}} - |w|_{\mathtt{3}}\big| \leq 1$. In particular, the letters $\mathtt{1}$ and $\mathtt{2}$ (respectively $\mathtt{0}$ and $\mathtt{3}$) alternate in $\mathbf{y}$.

Proof. First note that if for all factors of a word u, the numbers of two letters x and y differ by at most 1, then x and y alternate in u. Furthermore, if the first or the last occurrence of one of these letters is x, then $|u|_x \geq |u|_y$. If both the first and the last occurrences are x, then $|u|_x = |u|_y + 1$.

We prove the result by induction on the length ℓ of the factor. The result is true for factors of length $\ell = 1$. Let w be a factor of length $\ell > 1$ and assume the result holds for factors of length smaller than ℓ. If w can be de-substituted as $w = \nu(w')$, we have

$$\begin{aligned}
|w|_{\mathtt{0}} &= |w'|_{\mathtt{2}},\\
|w|_{\mathtt{1}} &= |w'|_{\mathtt{0}} + |w'|_{\mathtt{1}} + |w'|_{\mathtt{3}},\\
|w|_{\mathtt{2}} &= |w'|_{\mathtt{0}} + |w'|_{\mathtt{2}} + |w'|_{\mathtt{3}},\\
|w|_{\mathtt{3}} &= |w'|_{\mathtt{1}}.
\end{aligned}$$

Using the inductive hypothesis, we have

$$\big||w|_{\mathtt{1}} - |w|_{\mathtt{2}}\big| = \big||w|_{\mathtt{0}} - |w|_{\mathtt{3}}\big| = \big||w'|_{\mathtt{1}} - |w'|_{\mathtt{2}}\big| \leq 1.$$

If w cannot be de-substituted and has odd length, we have

$$w \in \big\{\mathtt{1}^{-1}\nu(w'),\, \mathtt{2}^{-1}\nu(w'),\, \nu(w')\mathtt{1},\, \nu(w')\mathtt{2}\big\}$$

for some factor w' with $|w'| < \ell$. Assume that $w = 1^{-1}\nu(w')$. Then as before, we have $\big||w|_0 - |w|_3\big| = \big||w'|_1 - |w'|_2\big| \leq 1$. For the numbers of 1 and 2, w' starts with 0 or 1. Since by Lemma 2.17 a 0 is always followed by a 1, w' starts either with 01 or with 1. In both cases, since 1 and 2 alternate, we have $|w'|_1 \geq |w'|_2$ and thus

$$\big||w|_1 - |w|_2\big| = \big||w'|_1 - |w'|_2 - 1\big| \leq 1.$$

The same reasoning can be done for $w = 2^{-1}\nu(w')$. If $w = \nu(w')1$, then we clearly have $\big||w|_0 - |w|_3\big| \leq 1$ using the result on $\nu(w')$. By Lemma 2.17, the factor $\nu(w')$ must end either with 0 or 2. So w' ends with 0 or 2 as well. Since a 0 is always preceded by a 2, we necessarily have $|w'|_2 \geq |w'|_1$ and

$$\big||w|_1 - |w|_2\big| = \big||w'|_1 - |w'|_2 + 1\big| \leq 1.$$

The same reasoning applies to $w = \nu(w')2$.

If w cannot be de-substituted and has even length, then we have

$$w \in \big\{1^{-1}\nu(w')1,\ 1^{-1}\nu(w')2,\ 2^{-1}\nu(w')1,\ 2^{-1}\nu(w')2\big\}$$

for some factor w' with $|w'| < \ell$. If the same letter is removed and added to $\nu(w')$, then the result is clearly true. Otherwise, assume that $w = 1^{-1}\nu(w')2$ (the same reasoning holds for the last case). It is clear that $\big||w|_0 - |w|_3\big| \leq 1$ using the result on $\nu(w')$. For the numbers of 1 and 2, as before, w' starts with 01 or 1 and ends with 13 or 1. Hence we have $|w'|_1 = |w'|_2 + 1$ and then

$$\big||w|_1 - |w|_2\big| = \big||w'|_1 - |w'|_2 - 2\big| \leq 1. \qquad \square$$

Lemma 2.19.

Let τ, τ' be the morphisms respectively defined by

$$\tau : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 2 \\ 2 \mapsto 1 \\ 3 \mapsto 3 \end{cases} \quad \text{and} \quad \tau' : \begin{cases} 0 \mapsto 3 \\ 1 \mapsto 1 \\ 2 \mapsto 2 \\ 3 \mapsto 0 \end{cases}.$$

If w is a factor of $\mathbf{y}$, then $\tau'(w)^{\mathrm{R}}$, $\tau(w)^{\mathrm{R}}$ and $\tau'(\tau(w))$ are also factors of $\mathbf{y}$.

Proof. We prove the lemma for $\tau'(w)^{\mathrm{R}}$ and $\tau(w)^{\mathrm{R}}$ since $\tau'(\tau(w)) = \tau'(\tau(w)^{\mathrm{R}})^{\mathrm{R}}$.

We first prove by induction that for any factor u starting with the letter x and ending with the letter y,

$$\tau'(\nu(u))^{\mathrm{R}} = a^{-1}\nu(\tau(u)^{\mathrm{R}})b \tag{2.3}$$

where $a = 1$ (respectively $a = 2$, $b = 1$, $b = 2$) if and only if $y \in \{0, 2\}$ (resp. $y \in \{1, 3\}$, $x \in \{0, 1\}$, $x \in \{2, 3\}$). Note that $a^{-1}\nu(\tau(u)^{\mathrm{R}})b$ is well defined. Indeed, if $y \in \{0, 2\}$, then $\tau(u)^{\mathrm{R}}$ starts with 0 or 1 and thus $\nu(\tau(u)^{\mathrm{R}})$ starts with $a = 1$. The same holds with $y \in \{1, 3\}$.

The relation (2.3) is true for u of length 1. We have for example

$$\tau'(\nu(0))^{\mathrm{R}} = 21 = 1^{-1}\nu(0)1 = 1^{-1}\nu(\tau(0)^{\mathrm{R}})1$$

and

$$\tau'(\nu(1))^{\mathrm{R}} = 01 = 2^{-1}\nu(2)1 = 2^{-1}\nu(\tau(1)^{\mathrm{R}})1.$$

Let $u = u'yx$ be a factor with at least two letters x and y. Assume the conclusion holds for words of length at most $|u| - 1$. By the inductive hypothesis, we have $\tau'(\nu(u'y))^{\mathrm{R}} = a^{-1}\nu(\tau(u'y)^{\mathrm{R}})b$ and $\tau'(\nu(x))^{\mathrm{R}} = c^{-1}\nu(\tau(x)^{\mathrm{R}})d$ with appropriate a, b, c, d. Since yx is a factor, one can check using Lemma 2.17 that $a = d$. Indeed, if $y \in \{0, 2\}$, then $x \in \{0, 1\}$. So $a = 1$ and $d = 1$. Similarly, if $y \in \{1, 3\}$, then $x \in \{2, 3\}$. Hence, $a = 2$ and $d = 2$. Thus, we have

$$\begin{aligned}
\tau'(\nu(u))^{\mathrm{R}} &= \tau'(\nu(u'yx)^{\mathrm{R}}) \\
&= \tau'(\nu(x))^{\mathrm{R}}\tau'(\nu(u'y))^{\mathrm{R}} \\
&= c^{-1}\nu(\tau(x)^{\mathrm{R}})da^{-1}\nu(\tau(u'y)^{\mathrm{R}})b \\
&= c^{-1}\nu(\tau(u'yx)^{\mathrm{R}})b \\
&= c^{-1}\nu(\tau(u)^{\mathrm{R}})b.
\end{aligned}$$

We can similarly prove by induction that for any factor u starting with the letter x and ending with the letter y,

$$\tau(\nu(u))^{\mathrm{R}} = a^{-1}\nu(\tau'(u)^{\mathrm{R}})b$$

where $a = 1$ (respectively $a = 2$, $b = 1$, $b = 2$) if and only if $y \in \{1, 3\}$ (resp. $y \in \{0, 2\}$, $x \in \{2, 3\}$, $x \in \{0, 1\}$).

We now prove the lemma (for τ and τ' together) by induction on the length of w. One can check by hand that the lemma is true for w of length at most 4. Assume the lemma is true for any factor of length at most $n \geq 4$, and let w be a factor of length $n+1$. There exist some factors s, t and v such that $swt = \nu(v)$, $0 \leq |t| \leq 1$ and $1 \leq |s| \leq 2$. Then we have $|v| \leq \frac{n+4}{2} \leq n$. By the inductive hypothesis, $\tau(v)^{\mathrm{R}}$ is a factor of $\mathbf{y}$. Hence $\nu(\tau(v)^{\mathrm{R}})$ is also a factor of $\mathbf{y}$. Using the previous result, $\tau'(\nu(v))^{\mathrm{R}} = a^{-1}\nu(\tau(v)^{\mathrm{R}})b$ for some letters a and b. But we also have $\tau'(\nu(v))^{\mathrm{R}} = \tau'(t)^{\mathrm{R}}\tau'(w)^{\mathrm{R}}\tau'(s)^{\mathrm{R}}$ and since s has at least one letter, $\tau'(w)^{\mathrm{R}}$ is a factor of $\nu(\tau(v)^{\mathrm{R}})$. Hence it is a factor of $\mathbf{y}$. We do the same proof for $\tau(w)^{\mathrm{R}}$. $\square$

We are now ready to prove the relationship between $\mathcal{P}_{\mathbf{y}}^{(1)}(n)$ and $\Delta_{12}(n)$.

Proof of Proposition 2.16. Let u be a factor of length n of $\mathbf{y}$. Let $n_{12} = |u|_1 + |u|_2$ and $n_{03} = |u|_0 + |u|_3$.

Assume first that n is odd. If n_{12} is even, then there are the same number of 1's and 2's in u by Lemma 2.18. Since n_{13} is odd, if $|u|_0 = |u|_3+1$ (resp. $|u|_3 = |u|_0+1$), then $\tau'(u)^{\mathrm{R}}$ is a factor by Lemma 2.19 and $|\tau'(u)^{\mathrm{R}}|_3 = |\tau'(u)^{\mathrm{R}}|_0+1$ (resp. $|\tau'(u)^{\mathrm{R}}|_0 = |\tau'(u)^{\mathrm{R}}|_3+1$). In either case, $\tau'(u)^{\mathrm{R}}$ still has n_{12} ones and twos. Hence there are exactly two abelian equivalence classes for fixed n odd and n_{12} even. We can do the same reasoning if n_{12} is odd. Finally, there are $\Delta_{12}(n)+1$ possible values for n_{12} and thus $2(\Delta_{12}(n)+1)$ abelian equivalence classes for a fixed odd n.

Assume now that n is even. If both n_{12} and n_{03} are even, then u necessarily has the same number of 1's as 2's and the same number of 0's as 3's, and thus there is only one abelian equivalence class. Hence assume that n_{12} and n_{03} are odd. We have $(|u|_0 - |u|_3, |u|_1 - |u|_2)$ in $\{-1, 1\}^2$. By Lemma 2.19, the four factors u, $\tau'(u)^{\mathrm{R}}$, $\tau(u)^{\mathrm{R}}$ and $\tau'(\tau(u))$ realize the four possibilities for $(|u|_0 - |u|_3, |u|_1 - |u|_2)$. Hence if n_{12} and n_{03} are both odd, there are four abelian equivalence classes.

Now, we just have to count pairs (n, n_{12}) with n and n_{12} even. If $\Delta_{12}(n)$ is odd, there are exactly $(\Delta_{12}(n)+1)/2$ such pairs. So there are

$$1 \cdot (\Delta_{12}(n)+1)/2 + 4 \cdot (\Delta_{12}(n)+1)/2 = \frac{5}{2}(\Delta_{12}(n)+1)$$

abelian classes for this value of n. If $\Delta_{12}(n)$ is even and $\min_{12}(n)$ is odd, there are exactly $\Delta_{12}(n)/2$ even values for n_{12}, and so there are

$$1 \cdot \Delta_{12}(n)/2 + 4 \cdot (\Delta_{12}(n)/2 + 1) = \frac{5}{2}\Delta_{12}(n) + 4$$

abelian classes. Finally, if $\Delta_{12}(n)$ is even and $\min_{12}(n)$ is even, there are exactly $\Delta_{12}(n)/2+1$ even values for n_{12}, and so there are

$$1 \cdot (\Delta_{12}(n)/2 + 1) + 4 \cdot \Delta_{12}(n)/2 = \frac{5}{2}\Delta_{12}(n) + 1$$

abelian classes. □

We now turn our attention to the sequences occurring in the expression of $\mathcal{P}_{\mathbf{y}}^{(1)}$ derived in Proposition 2.16.

Proposition 2.20.

Let $\ell \geq 1$ and $0 \leq r < 2^{\ell}$. We have

$$\Delta_{12}(2^{\ell}+r) = \begin{cases} \Delta_{12}(r)+1 & \text{if } r \leq 2^{\ell-1} \\ \Delta_{12}(2^{\ell+1}-r) & \text{if } r > 2^{\ell-1}. \end{cases}$$

Moreover,

$$\min_{12}(2^{\ell}+r) \equiv \begin{cases} \min_{12}(r)+\ell \pmod 2 & \text{if } r \leq 2^{\ell-1} \\ \min_{12}(2^{\ell+1}-r) + \Delta_{12}(2^{\ell+1}-r) \pmod 2 & \text{otherwise.} \end{cases}$$

Note that those latter relations have a form similar to (but slightly different from) the assumptions of Theorem 2.11. Before giving the proof, we prove a corollary. The 2-regularity of $\mathcal{P}_{\mathbf{y}}^{(1)}(n)_{n\geq 0}$ follows from Proposition 2.16 and Corollary 2.21.

Corollary 2.21.

- The sequence $\Delta_{12}(n)_{n\geq 0}$ is 2-regular.
- The sequence $(\Delta_{12}(n) \bmod 2)_{n\geq 0}$ is 2-automatic.
- The sequence $(\min_{12}(n) \bmod 2)_{n\geq 0}$ is 2-automatic.

Proof. The first assertion is a direct consequence of Proposition 2.20 and Theorem 2.11. The second assertion follows from Lemma 1.31.

To prove the last assertion, we prove by induction that, modulo 2,

$$\min_{12}(16n+i) \equiv \begin{cases} \min_{12}(4n) & \text{if } i = 0 \\ \min_{12}(4n+1) & \text{if } i \in \{1,4,5\} \\ \min_{12}(4n+1)+1 & \text{if } i \in \{2,3\} \\ \min_{12}(4n+2) & \text{if } i \in \{6,8,9\} \\ \min_{12}(4n+2)+1 & \text{if } i \in \{7,10\} \\ \min_{12}(4n+3) & \text{if } i \in \{12,13,15\} \\ \min_{12}(4n+3)+1 & \text{if } i \in \{11,14\} \end{cases}$$

and

$$\Delta_{12}(16n+i) \equiv \begin{cases} \Delta_{12}(4n) & \text{if } i = 0 \\ \Delta_{12}(4n+1) & \text{if } i \in \{1,2,4\} \\ \Delta_{12}(4n+1)+1 & \text{if } i \in \{3,5\} \\ \Delta_{12}(4n+2) & \text{if } i = 8 \\ \Delta_{12}(4n+2)+1 & \text{if } i \in \{6,7,9,10\} \\ \Delta_{12}(4n+3) & \text{if } i \in \{12,14,15\} \\ \Delta_{12}(4n+3)+1 & \text{if } i \in \{11,13\}. \end{cases}$$

The relations are true for $n = 0$. Let $n > 0$ and assume they are true for $n' < n$. We can write $n = 2^\ell + r$ with $\ell \geq 0$ and $0 \leq r < 2^\ell$. Let $i \in \{0, \ldots, 15\}$. We consider two cases.

Assume first that $r < 2^{\ell-1}$. We have $16n + i = 2^{\ell+4} + 16r + i$ and $16r + i < 2^{\ell+3}$.

$$\begin{aligned} \min_{12}(16n+i) &\equiv \min_{12}(16r+i) + \ell + 4 && \text{(by Proposition 2.20)} \\ &\equiv \min_{12}(4r+j) + \delta + \ell + 4 && \text{(by inductive hypothesis)} \\ &\equiv \min_{12}(2^{\ell+2} + 4r + j) + \delta && \text{(by Proposition 2.20)} \\ &\equiv \min_{12}(4n+j) + \delta \pmod 2 \end{aligned}$$

for some $j \in \{0, \ldots, 3\}$ and $\delta \in \{0, 1\}$ according to the relations. A similar reasoning holds for the Δ_{12} relations.

Assume now that $r \geq 2^{\ell-1}$ and $i \neq 0$. Setting $i' = 16 - i$ and $n' = 2^{\ell+1} - r - 1$, we obtain $16n' + i' = 2^{\ell+5} - 16r - i$. It follows that, by Proposition 2.20,

$$\begin{aligned} \min_{12}(16n+i) &\equiv \min_{12}(2^{\ell+5} - 16r - i) + \Delta_{12}(2^{\ell+5} - 16r - i) \\ &\equiv \min_{12}(16n' + i') + \Delta_{12}(16n' + i') \\ &\equiv \min_{12}(4n' + k) + \delta + \Delta_{12}(4n' + k') + \delta' && \text{(by inductive hypothesis)} \end{aligned}$$

for some $k, k' \in \{0, \ldots, 3\}$ and $\delta, \delta' \in \{0, 1\}$ according to the relations. Note that we have $k = k'$, so

$$\begin{aligned} \min_{12}(16n+i) &\equiv \min_{12}(4n' + k) + \delta + \Delta_{12}(4n' + k) + \delta' \\ &\equiv \min_{12}(2^{\ell+3} - (4r + 4 - k)) + \delta + \Delta_{12}(2^{\ell+3} - (4r + 4 - k)) + \delta' \\ &\equiv \min_{12}(2^{\ell+2} + (4r + 4 - k)) + \delta + \delta' && \text{(by Proposition 2.20)} \\ &\equiv \min_{12}(4n + (4 - k)) + \delta + \delta' \pmod 2. \end{aligned}$$

Table 2.2 gives the values of i', k, δ and δ' for all the values of $i \neq 0$. Observe that the values of $4-k$ and ($\delta + \delta'$ mod 2) are the values given in the relation for i. To conclude the proof, consider the case $i = 0$. We have

$$\begin{aligned}
\min_{12}(16n) &\equiv \min_{12}(16(2^{\ell+1}-r)) + \Delta_{12}(16(2^{\ell+1}-r)) && \text{(Proposition 2.20)}\\
&\equiv \min_{12}(4(2^{\ell+1}-r)) + \Delta_{12}(4(2^{\ell+1}-r)) && \text{(by inductive hypothesis)}\\
&\equiv \min_{12}(4n) \pmod 2 && \text{(Proposition 2.20).}
\end{aligned}$$

A similar reasoning works for the Δ_{12} relations. □

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
i'	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
k	3	3	3	3	3	2	2	2	2	2	1	1	1	1	1
δ	0	1	0	0	1	1	0	0	1	0	0	0	1	1	0
δ'	0	0	1	0	1	1	1	0	1	1	1	0	1	0	0

Table 2.2: The corresponding values of $i' = 16 - i$, k, δ and δ'.

Proposition 2.20 is a direct consequence of Lemmas 2.22, 2.25 and 2.27 given below.

Lemma 2.22.

Let $\ell \in \mathbb{N}$, $\ell \geq 1$. We have $\Delta_{12}(2^\ell) = 1$, $\min_{12}(2^\ell) \equiv \ell \pmod 2$,

$$\min_{12}(2^\ell) + \max_{12}(2^{\ell+1}) = 2^{\ell+1} \text{ and } \max_{12}(2^\ell) + \min_{12}(2^{\ell+1}) = 2^{\ell+1}.$$

Proof. Let $\ell \geq 1$, $A_\ell = \frac{2^{\ell+1}+(-1)^\ell}{3}$ and $B_\ell = \frac{2^{\ell+1}+2(-1)^{\ell+1}}{3}$. The sequences

$$(A_\ell)_{\ell\geq 1} = (1,3,5,11,21,\ldots) \text{ and } (B_\ell)_{\ell \geq 1} = (2,2,6,10,22,\ldots)$$

are integer sequences and both satisfy the recurrence relation $X_{\ell+1} = 2^{\ell+1} - X_\ell$. Moreover we have $A_\ell = B_\ell + 1$ for even ℓ and $B_\ell = A_\ell + 1$ for odd ℓ. Note that $|\nu^\ell(\mathtt{1})|_{\mathtt{1}} + |\nu^\ell(\mathtt{1})|_{\mathtt{2}} = A_\ell$ and $|\nu^\ell(\mathtt{0})|_{\mathtt{1}} + |\nu^\ell(\mathtt{0})|_{\mathtt{2}} = B_\ell$.

We show by induction that

$$\left\{|w|_{\mathtt{1}} + |w|_{\mathtt{2}} : w \text{ factor of } \mathbf{y} \text{ with } |w| = 2^\ell\right\} = \{A_\ell, B_\ell\}.$$

Note that this result will imply the lemma and that we already have A_ℓ and B_ℓ in the set.

It is easy to check the result for $\ell = 1$. Assume the result is true for $\ell \geq 1$. Let w be a factor of $\mathbf{y}$ of length $2^{\ell+1}$. If w can be de-substituted, then $w = \nu(u)$ and we have $|w|_{\mathtt{1}} + |w|_{\mathtt{2}} = 2|u|_{\mathtt{0}} + |u|_{\mathtt{1}} + |u|_{\mathtt{2}} + 2|u|_{\mathtt{3}}$ as in the proof of Lemma 2.18. Hence we obtain $|w|_{\mathtt{1}} + |w|_{\mathtt{2}} = 2|u| - (|u|_{\mathtt{1}} + |u|_{\mathtt{2}}) = 2^{\ell+1} - (|u|_{\mathtt{1}} + |u|_{\mathtt{2}})$. Using the recurrence relation for A_ℓ and B_ℓ and since $|u|_{\mathtt{1}} + |u|_{\mathtt{2}} \in \{A_\ell, B_\ell\}$, we have $|w|_{\mathtt{1}} + |w|_{\mathtt{2}} \in \{A_{\ell+1}, B_{\ell+1}\}$. If w cannot be de-substituted, then we can write $w = a^{-1}\nu(u)b$ for some letters $a, b \in \{\mathtt{1}, \mathtt{2}\}$ and $|\nu(u)| = 2^{\ell+1}$. So $|w|_{\mathtt{1}} + |w|_{\mathtt{2}} = |\nu(u)|_{\mathtt{1}} + |\nu(u)|_{\mathtt{2}}$. Since we already proved that $|\nu(u)|_{\mathtt{1}} + |\nu(u)|_{\mathtt{2}}$ is in $\{A_{\ell+1}, B_{\ell+1}\}$, we are done.

To prove the second assertion of the lemma, observe that $\min_{12}(2^\ell) = A_\ell$ if ℓ is odd and $\min_{12}(2^\ell) = B_\ell$ if ℓ is even. Furthermore, A_ℓ is always odd whereas B_ℓ is always even. □

In order to prove Lemmas 2.25 and 2.27, we first need some technical results.

Lemma 2.23.

Let u be a factor of $\mathbf{y}$ of length n. We have

- $|u|_1 + |u|_{2=} \max_{12}(n)$ if and only if $|\nu(u)|_1 + |\nu(u)|_2 = \min_{12}(2n)$,
- $|u|_1 + |u|_2 = \min_{12}(n)$ if and only if $|\nu(u)|_1 + |\nu(u)|_2 = \max_{12}(2n)$.

Proof. Recall that $|\nu(u)|_1 + |\nu(u)|_2 = 2n - (|u|_1 + |u|_2)$. Assume that $|u|_1 + |u|_2 = \max_{12}(n)$ and that $|\nu(u)|_1 + |\nu(u)|_2 = x > \min_{12}(2n)$. Thus $x = 2n - \max_{12}(n)$. There exists a factor w of length $2n$ with $x - 1$ ones and twos. We can assume that w can be de-substituted. Otherwise, we can write w as $w = a^{-1}\nu(v)b$ for some $a, b \in \{1, 2\}$. Thus $\nu(v)$ has the same length as w and the same number of 1's and 2's. So we can assume $w = \nu(v)$. Then $|v|_1 + |v|_2 = 2n - (x - 1) = \max_{12}(n) + 1$, a contradiction.

For the other direction, assume that $|u|_1+|u|_2 = x < \max_{12}(n)$ and that $|\nu(u)|_1+|\nu(u)|_2 = \min_{12}(2n)$. Thus $x = n - \min_{12}(n)$. As before, there exists a factor v of length n with $x + 1$ ones and twos. Then $\nu(v)$ has $\min_{12}(n) - 1$ ones and twos, a contradiction.

The second part of the lemma is similar. □

Lemma 2.24.

Let n be an odd integer. Then we have

$$\begin{aligned}\min_{12}(n) &= \min_{12}(n + 1) - 1,\\ \max_{12}(n) &= \max_{12}(n - 1) + 1.\end{aligned}$$

Proof. Let u be a factor of even length $n + 1$ minimizing the number of 1's and 2's. Then either u starts with 1 or 2, or ends with 1 or 2. Indeed, if u can be de-substituted, then it starts with 1 or 2. Otherwise, its last letter is the beginning of an image of ν and thus is 1 or 2. Removing this letter, we get a word of length n with $\min_{12}(n+1)-1$ ones and twos. Since the function $\min_{12}$ increases by 0 or 1 from n to $n+1$, we have $\min_{12}(n) = \min_{12}(n+1) - 1$.

For the second equality, consider a factor u of even length $n - 1$ with $\max_{12}(n - 1)$ ones and twos. There exist two letters a and b such that aub is a factor. Then, as before, since aub has even length, a or b must be a 1 or a 2. Then au or ub is a factor of length n with $\max_{12}(n - 1) + 1$ ones and twos and we conclude as before. □

Lemma 2.25.

If $\ell \geq 1$ and $0 \leq r \leq 2^{\ell-1}$, then

$$\begin{aligned}\max_{12}(2^\ell + r) &= \max_{12}(2^\ell) + \max_{12}(r)\\ \min_{12}(2^\ell + r) &= \min_{12}(2^\ell) + \min_{12}(r).\end{aligned}$$

Proof. We prove the two results together by induction on ℓ. One checks the case $\ell = 1$. Let $\ell > 1$ and assume the result is true for $\ell - 1$. Let $0 \leq r \leq 2^{\ell-1}$.

Assume first that r is even. By the inductive hypothesis, there exists a factor u of length $2^{\ell-1} + r/2$ such that

$$|u|_1 + |u|_2 = \min_{12}(2^{\ell-1} + r/2) = \min_{12}(2^{\ell-1}) + \min_{12}(r/2).$$

We can write $u = vw$ with v of length $2^{\ell-1}$ and w of length $r/2$. Both the words v and w must minimize the number of 1's and 2's for their respective lengths. By Lemma 2.23, $\nu(u) = \nu(v)\nu(w)$ maximizes the number of 1's and 2's and so do $\nu(v)$ and $\nu(w)$. Thus, $\max_{12}(2^\ell + r) = |\nu(u)|_1 + |\nu(u)|_2$ and

$$\max_{12}(2^\ell + r) = |\nu(v)|_1 + |\nu(v)|_2 + |\nu(w)|_1 + |\nu(w)|_2 = \max_{12}(2^\ell) + \max_{12}(r).$$

A similar proof shows that $\min_{12}(2^\ell + r) = \min_{12}(2^\ell) + \min_{12}(r)$.

Assume now that r is odd. We still have $0 \leq r - 1 < r + 1 \leq 2^{\ell-1}$. Hence we can apply the previous result to obtain $\max_{12}(2^\ell + r - 1) = \max_{12}(2^\ell) + \max_{12}(r - 1)$. By Lemma 2.24,

$$\begin{aligned} \max_{12}(2^\ell + r) &= \max_{12}(2^\ell + r - 1) + 1 \\ &= \max_{12}(2^\ell) + \max_{12}(r - 1) + 1 \\ &= \max_{12}(2^\ell) + \max_{12}(r). \end{aligned}$$

For the $\min_{12}$ equality, a similar argument holds (using the previous result for $r + 1$). □

Lemma 2.26.

If $\ell \geq 1$ and $2^{\ell-1} \leq r \leq 2^\ell$, then

$$\begin{aligned} \max_{12}(2^{\ell+1}) &= \max_{12}(2^\ell + r) + \min_{12}(2^\ell - r) \\ \min_{12}(2^{\ell+1}) &= \min_{12}(2^\ell + r) + \max_{12}(2^\ell - r). \end{aligned}$$

Moreover, there is a factor of length $2^{\ell+1}$ maximizing (resp. minimizing) the number of 1's and 2's such that the prefix of length $2^\ell + r$ also maximizes (resp. minimizes) the number of 1's and 2's.

Proof. We proceed by induction on ℓ. The result is true for $\ell = 1$ since the only non-trivial case is $r = 1$. Then $\max_{12}(4) = \max_{12}(3) + \min_{12}(1)$ and $\min_{12}(4) = \min_{12}(3) + \max_{12}(1)$ and the factors 2120 and 0132 satisfy the claim.

Let $\ell > 1$ and assume the result is true for $\ell - 1$. Let $2^{\ell-1} \leq r \leq 2^\ell$. Assume first that r is even. Then $2^{\ell-2} \leq r/2 \leq 2^{\ell-1}$. By the inductive hypothesis, there is a factor u of length 2^ℓ minimizing the number of 1's and 2's such that the prefix v of length $2^{\ell-1} + r/2$ minimizes the number of 1's and 2's. Thus we can write $u = vw$ and $|v|_1 + |v|_2 = \min_{12}(2^{\ell-1} + r/2)$ and necessarily $|w|_1 + |w|_2 = \max_{12}(2^{\ell-1} - r/2)$. By Lemma 2.23, $\nu(u)$ and $\nu(v)$ maximize the number of 1's and 2's and $\nu(w)$ minimizes the number of 1's and 2's. So we can conclude the result. A similar proof shows the other relation. If r is odd, then we still have $2^{\ell-1} \leq r-1 \leq 2^\ell$ since $\ell > 1$. Thus we can use the previous result and together with Lemma 2.24, we have

$$\begin{aligned} \max_{12}(2^{\ell+1}) &= \max_{12}(2^\ell + r - 1) + \min_{12}(2^\ell - r + 1) \\ &= \max_{12}(2^\ell + r) - 1 + \min_{12}(2^\ell - r) + 1 \\ &= \max_{12}(2^\ell + r) + \min_{12}(2^\ell - r). \end{aligned}$$

Similarly, using the fact that $r+1 \leq 2^\ell$,

$$\begin{aligned}\min\nolimits_{12}(2^{\ell+1}) &= \min\nolimits_{12}(2^\ell + r + 1) + \max\nolimits_{12}(2^\ell - r - 1)\\ &= \min\nolimits_{12}(2^\ell + r) + 1 + \max\nolimits_{12}(2^\ell - r) - 1\\ &= \min\nolimits_{12}(2^\ell + r) + \max\nolimits_{12}(2^\ell - r).\end{aligned}$$

For the construction of the factors, one can construct them using the factor $\nu(u)$ maximizing the number of $\mathtt{1}$'s and $\mathtt{2}$'s given for $r-1$ and the factor $\nu(u')$ minimizing the number of $\mathtt{1}$'s and $\mathtt{2}$'s given for $r+1$ in the previous construction. Since r is odd, the letter between the prefix $\nu(v)$ of length $2^\ell + r - 1$ and $2^\ell + r$ of $\nu(u)$ is 1 or 2. Since the prefix of length $2^\ell + r - 1$ of $\nu(u)$ maximizes the number of $\mathtt{1}$'s and $\mathtt{2}$'s, so does the prefix of length $2^\ell + r$ of $\nu(u)$. For $\min_{12}$, consider $\nu(u')$. There exist letters a and b such that $w = a^{-1}\nu(u')b$ is still a factor. We must have $a, b \in \{\mathtt{1}, \mathtt{2}\}$. Then the prefix of length $2^\ell + r$ of w minimizes the number of $\mathtt{1}$'s and $\mathtt{2}$'s. ☐

The previous lemma permits us to reformulate some relations between the two sequences $\max_{12}(n)_{n\geq 0}$ and $\min_{12}(n)_{n\geq 0}$.

Lemma 2.27.

If $\ell \geq 1$ and $2^{\ell-1} \leq r \leq 2^\ell$, then

$$\begin{aligned}\max\nolimits_{12}(2^\ell + r) &= 2^{\ell+1} - \min\nolimits_{12}(2^{\ell+1} - r)\\ \min\nolimits_{12}(2^\ell + r) &= 2^{\ell+1} - \max\nolimits_{12}(2^{\ell+1} - r).\end{aligned}$$

Proof. From the previous lemma, we have

$$\max\nolimits_{12}(2^\ell + r) = \max\nolimits_{12}(2^{\ell+1}) - \min\nolimits_{12}(2^\ell - r).$$

Note that, by Lemma 2.22, we have $\max_{12}(2^{\ell+1}) = 2^{\ell+1} - \min_{12}(2^\ell)$. Moreover, we get by Lemma 2.25

$$\min\nolimits_{12}(2^\ell - r) = \min\nolimits_{12}(2^\ell + 2^\ell - r) - \min\nolimits_{12}(2^\ell),$$

since $0 \leq 2^\ell - r \leq 2^{\ell-1}$. Similar relations hold when changing $\max_{12}$ to $\min_{12}$. ☐

The proof of Proposition 2.20 about the reflection relation satisfied by $\Delta_{12}(n)$ and the recurrence relation of $\min_{12}(n)$ is now immediate.

Proof of Proposition 2.20. If $\ell \geq 1$ and $0 \leq r \leq 2^{\ell-1}$, then subtracting the two relations provided by Lemma 2.25 gives

$$\Delta_{12}(2^\ell + r) = \Delta_{12}(\ell) + \Delta_{12}(r)$$

and we can conclude using the first relation given in Lemma 2.22, $\Delta_{12}(2^\ell) = 1$. By Lemma 2.25, $\min_{12}(2^\ell + r) \equiv \min_{12}(2^\ell) + \min_{12}(r) \pmod 2$. The expression for $\min_{12}(2^\ell + r)$ follows since $\min_{12}(2^\ell) \equiv \ell \pmod 2$ by Lemma 2.22.

If $\ell \geq 1$ and $2^{\ell-1} < r < 2^\ell$, then subtracting the two relations provided by Lemma 2.27 permits us to conclude the expression claimed for $\Delta_{12}(2^\ell + r)$. Moreover, using Lemma 2.27, we get

$$\begin{aligned}\min\nolimits_{12}(2^\ell + r) &\equiv \max\nolimits_{12}(2^{\ell+1} - r) \pmod 2 \\ &\equiv \min\nolimits_{12}(2^{\ell+1} - r) + \Delta_{12}(2^{\ell+1} - r) \pmod 2. \qquad \square\end{aligned}$$

Using Propositions 2.16 and 2.20, we can express recurrence relations for the abelian complexity $\mathcal{P}_{\mathbf{y}}^{(1)}$ of the 2-block coding $\mathbf{y}$ of the Thue–Morse word. These recurrence relations are similar to the ones of Theorem 2.11; unfortunately they do not completely coincide. Hence, we cannot deduce immediately the 2-regularity of $\mathcal{P}_{\mathbf{y}}^{(1)}$ from these relations.

Theorem 2.28.

Let $\ell \geq 2$ and $0 \leq r < 2^\ell$. For $r \leq 2^{\ell-1}$, we have

$$\mathcal{P}_{\mathbf{y}}^{(1)}(2^\ell + r) = \begin{cases} \mathcal{P}_{\mathbf{y}}^{(1)}(r) + 2 & \text{if } r \text{ is odd} \\ \mathcal{P}_{\mathbf{y}}^{(1)}(r) + 1 & \text{if } (r,\ \Delta_{12}(2^\ell + r) \text{ and } \min_{12}(2^\ell + r) \text{ are even}) \\ & \text{or } (r \text{ and } \Delta_{12}(2^\ell + r) + 1 \text{ are even} \\ & \quad \text{and } \min_{12}(2^\ell + r) \equiv \ell + 1 \pmod 2) \\ \mathcal{P}_{\mathbf{y}}^{(1)}(r) + 4 & \text{otherwise.} \end{cases}$$

For $r > 2^{\ell-1}$, we have $\mathcal{P}_{\mathbf{y}}^{(1)}(2^\ell + r) = \mathcal{P}_{\mathbf{y}}^{(1)}(2^{\ell+1} - r)$.

2.3.2 The 2-abelian complexity of t is 2-regular

To prove the 2-regularity of $\mathcal{P}_{\mathbf{t}}^{(2)}$, the aim of this subsection is to express $\mathcal{P}_{\mathbf{t}}^{(2)}(n+1)$ in terms of $\mathcal{P}_{\mathbf{y}}^{(1)}(n)$, $\Delta_{12}(n)$, $(\min_{12}(n) \bmod 2)$ and two new functions $\mathrm{MJ}_{03}(n)$ and $\mathrm{mj}_{03}(n)$.

Let

$$\begin{aligned}\max\nolimits_{03}(n) &:= \max\{|u|_0 + |u|_3 \ : \ u \text{ is a factor of } \mathbf{y} \text{ with } |u| = n\}, \\ \min\nolimits_{03}(n) &:= \min\{|u|_0 + |u|_3 \ : \ u \text{ is a factor of } \mathbf{y} \text{ with } |u| = n\},\end{aligned}$$

We define the *max-jump* function $\mathrm{MJ}_{03}(n) : \mathbb{N} \to \{0,1\}$ by $\mathrm{MJ}_{03}(0) = 0$ and for $n \geq 1$,

$$\mathrm{MJ}_{03}(n) := \begin{cases} 1 & \text{if } \max_{03}(n) > \max_{03}(n-1) \\ 0 & \text{otherwise.} \end{cases}$$

In other words, $\mathrm{MJ}_{03}(n) = 1$ when the function $\max_{03}$ increases. Similarly, the *min-jump* function $\mathrm{mj}_{03}(n) : \mathbb{N} \to \{0,1\}$ is defined by

$$\mathrm{mj}_{03}(n) := \begin{cases} 1 & \text{if } \min_{03}(n+1) > \min_{03}(n) \\ 0 & \text{otherwise.} \end{cases}$$

Remark 2.29. Observe that we could define the max-jump and min-jump functions using the functions $\max_{12}(n)$ and $\min_{12}(n)$ since we have $\max_{03}(n) = n - \min_{12}(n)$ and $\min_{03}(n) = n - \max_{12}(n)$. So we have for all $n \geq 0$

$$\mathrm{MJ}_{03}(n+1) = \min\nolimits_{12}(n) - \min\nolimits_{12}(n+1) + 1$$

and

$$\begin{aligned}\mathrm{mj}_{03}(n) &= \max_{12}(n) - \max_{12}(n+1) + 1\\ &= \min_{12}(n) - \min_{12}(n+1) + \Delta_{12}(n) - \Delta_{12}(n+1) + 1.\end{aligned}$$

We chose to define these two new functions with $\max_{03}(n)$ and $\min_{03}(n)$ to obtain similar results in the case of the 2-abelian complexity of the period-doubling word (Section 2.4).

The relationship between these functions and $\mathcal{P}^{(2)}_{\mathbf{t}}$ and $\mathcal{P}^{(1)}_{\mathbf{y}}$ is given in the following result.

Theorem 2.30.

Let $n \in \mathbb{N}$. If n is odd, the difference $\mathcal{P}^{(2)}_{\mathbf{t}}(n+1) - \mathcal{P}^{(1)}_{\mathbf{y}}(n)$ is equal to

$$\begin{cases} \Delta_{12}(n) + 2 - 2\,\mathrm{MJ}_{03}(n) - 2\,\mathrm{mj}_{03}(n) & \text{if } \min_{12}(n) \text{ and } \Delta_{12}(n) \text{ are even}\\ \Delta_{12}(n) + 1 - 2\,\mathrm{MJ}_{03}(n) & \text{if } \min_{12}(n) \text{ and } \Delta_{12}(n)+1 \text{ are even}\\ \Delta_{12}(n) + 1 - 2\,\mathrm{mj}_{03}(n) & \text{if } \min_{12}(n) \text{ and } \Delta_{12}(n) \text{ are odd}\\ \Delta_{12}(n) & \text{if } \min_{12}(n)+1 \text{ and } \Delta_{12}(n) \text{ are even.}\end{cases}$$

For n even, the difference $\mathcal{P}^{(2)}_{\mathbf{t}}(n+1) - \mathcal{P}^{(1)}_{\mathbf{y}}(n)$ is

$$\begin{cases} \frac{1}{2}\Delta_{12}(n) + 1 & \text{if } \min_{12}(n) \text{ and } \Delta_{12}(n) \text{ are even}\\ \frac{1}{2}\Delta_{12}(n) & \text{if } \min_{12}(n)+1 \text{ and } \Delta_{12}(n) \text{ are even}\\ \frac{1}{2}\Delta_{12}(n) + \frac{1}{2} & \text{if } \Delta_{12}(n) \text{ is odd.}\end{cases}$$

We require several preliminary results.

Proposition 2.31.

Let u and v be factors of $\mathbf{t}$ of length n. Let u' and v' be the 2-block codings of u and v. The factors u and v are 2-abelian equivalent if and only if u' and v' (of length $n-1$) are abelian equivalent and either u' and v' both have first letter in $\{\mathtt{0},\mathtt{1}\}$ or both have first letter in $\{\mathtt{2},\mathtt{3}\}$.

To compute $\mathcal{P}^{(2)}_{\mathbf{t}}$, we will use the abelian complexity of $\mathbf{y} = \mathrm{block}(\mathbf{t},2)$, $\mathcal{P}^{(1)}_{\mathbf{t}}$, and study when an abelian equivalence class of length-n factors of $\mathbf{y}$ splits into two 2-abelian equivalence classes of factors of length $n+1$ of $\mathbf{t}$. In other words, we study when two abelian equivalent factors of $\mathbf{y}$ can start, respectively, with a letter in $\{\mathtt{0},\mathtt{1}\}$ and with a letter in $\{\mathtt{2},\mathtt{3}\}$.

Let $\mathcal{X}$ be an abelian equivalence class of factors of $\mathbf{y}$ of length n. For a letter a, let n_a denote the number of a's in each element of $\mathcal{X}$ and let $n_{12} = n_{\mathtt{1}} + n_{\mathtt{2}}$, $n_{03} = n_{\mathtt{0}} + n_{\mathtt{3}}$.

Lemma 2.32.

If n_{12} is odd, then $\mathcal{X}$ leads to a unique 2-abelian equivalence class of $\mathbf{t}$.

Proof. Assume that $n_1 > n_2$ (the other case is similar). Then a word of $\mathcal{X}$ cannot start with $\mathtt{2}$ since the letters $\mathtt{1}$ and $\mathtt{2}$ alternate in $\mathbf{y}$ by Lemma 2.18. It cannot start with $\mathtt{3}$ neither since $n_1 > n_2$ and a $\mathtt{3}$ is always followed by $\mathtt{2}$ by Lemma 2.17. Hence it starts with $\mathtt{0}$ or $\mathtt{1}$. Thus $\mathcal{X}$ leads to a unique 2-abelian equivalence class. □

Lemma 2.33.

If n and n_{12} are even, then $\mathcal{X}$ splits into two 2-abelian equivalence classes of $\mathbf{t}$.

Proof. If n and n_{12} are even, then n_{03} is also even and thus $n_1 = n_2$ and $n_0 = n_3$. Let u be an element of $\mathcal{X}$. Then $u' = \tau'(\tau(u))$ is also an element of $\mathcal{X}$. Moreover, the first letter of u is in $\{\mathtt{0}, \mathtt{1}\}$ if and only if the first letter of u' is in $\{\mathtt{2}, \mathtt{3}\}$. Hence $\mathcal{X}$ splits into two 2-abelian equivalence classes. □

So the last and hardest case happens when n is odd and n_{12} is even, i.e., when n and n_{03} are odd. The MJ_{03} and mj_{03} functions permit us to handle this case.

Lemma 2.34.

Let n and n_{03} are odd. Let $a \in \{\mathtt{0}, \mathtt{3}\}$ (resp. $b \in \{\mathtt{0}, \mathtt{3}\}$) be the letter in majority (resp. in minority) in factors in $\mathcal{X}$, among $\{\mathtt{0}, \mathtt{3}\}$.

- We have $n_{03} = \max_{03}(n)$ and $\mathrm{MJ}_{03}(n) = 1$ if and only if every factor in $\mathcal{X}$ starts and ends with a.
- We have $n_{03} = \min_{03}(n)$ and $\mathrm{mj}_{03}(n) = 1$ if and only if every factor in $\mathcal{X}$ is preceded and followed by b.

Proof. Assume that $a = \mathtt{0}$ and $b = \mathtt{3}$ (the other case is symmetric). We first prove the statement for the maximum. Assume that all the factors in $\mathcal{X}$ start and end with $\mathtt{0}$. If $n_{03} < \max_{03}(n)$, by continuity of the number of $\mathtt{0}$'s and $\mathtt{3}$'s and since $\mathbf{y}$ is uniformly recurrent, there exists a factor yuz such that the factor yu (resp. uz) is of length n with n_{03} (resp. $n_{03}+1$) zeros and threes. We necessarily have $z \in \{\mathtt{0}, \mathtt{3}\}$ and u is not finishing with a letter in $\{\mathtt{0}, \mathtt{3}\}$. Since yu has n_{03} zeros and threes, yu or $\tau'(yu)^{\mathrm{R}}$ is an element of $\mathcal{X}$ that is either not finishing or not starting with $\mathtt{0}$, a contradiction. Hence we have $n_{03} = \max_{03}(n)$. Assume now that $\max_{03}(n-1) = n_{03}$. There exists a factor u of even length $n-1$ with n_{03} zeros. Without loss of generality, we can assume that u has more $\mathtt{0}$'s than $\mathtt{3}$'s (otherwise one can consider $\tau'(u)^{\mathrm{R}}$ by Lemma 2.19). Since u has even length, either u occurs at an even index in $\mathbf{y}$ and is always followed by $\mathtt{1}$ or $\mathtt{2}$, or u occurs at an odd index in $\mathbf{y}$ and is always preceded by $\mathtt{1}$ or $\mathtt{2}$. In other words, there is a factor of the form yu or uy with $y \in \{\mathtt{1}, \mathtt{2}\}$. Then yu or uy is an element of $\mathcal{X}$ with the first or last letter different from $\mathtt{0}$, a contradiction.

For the other direction, assume that $n_{03} = \max_{03}(n)$ and $\mathrm{MJ}_{03}(n) = 1$. Let u be a factor in $\mathcal{X}$. If $u = xu'$ or $u = u'x$ with a letter $x \neq \mathtt{0}$, then u' has length $n-1$ and n_{03} zeros and threes. Thus $\mathrm{MJ}_{03}(n) = 0$, a contradiction.

The second statement is proved in the same way. Assume that all the factors in $\mathcal{X}$ are preceded and followed by $\mathtt{3}$. If $n_{03} > \min_{03}(n)$, by continuity of the number of $\mathtt{0}$'s and $\mathtt{3}$'s and since $\mathbf{y}$ is uniformly recurrent, there exists a factor yuz such that the factor yu (resp. uz) is of length n with n_{03} (resp. $n_{03}-1$) zeros and threes. We necessarily have $z \in \{\mathtt{1}, \mathtt{2}\}$. Then as before yu or $\tau'(yu)^{\mathrm{R}}$ is and element of $\mathcal{X}$ that is either not always followed or not always preceded by $\mathtt{3}$, a contradiction. Hence we have $n_{03} = \min_{03}(n)$. Assume now that $\min_{03}(n+1) = n_{03}$. There exists a factor u of even length $n+1$ with n_{03} zeros. Without loss of generality, we can assume that u has more $\mathtt{0}$'s than $\mathtt{3}$'s (otherwise one can consider $\tau'(u)^{\mathrm{R}}$ by Lemma 2.19). Since u has even length, either u occurs at an even index and starts with $\mathtt{1}$ or $\mathtt{2}$ or u occurs at an odd index and ends with $\mathtt{1}$ or $\mathtt{2}$. In other words, $u = yu'$ or $u = u'y$ with $y \in \{\mathtt{1}, \mathtt{2}\}$ and u' is an element of $\mathcal{X}$ preceded or followed by a letter different from $\mathtt{3}$, a contradiction.

For the other direction, assume that $n_{03} = \min_{03}(n)$ and $\mathrm{mj}_{03}(n) = 1$. Let u be a factor in $\mathcal{X}$. If $u' = ux$ or $u' = xu$ is a factor with $x \in \{\mathtt{1}, \mathtt{2}\}$, then u' has length $n+1$ and n_{03} zeros and threes. So $\mathrm{mj}_{03}(n) = 0$, which is a contradiction. Observe also that it is impossible to have $\mathtt{0}u$ or $u\mathtt{0}$ as factors of $\mathbf{y}$ since $|u|_{\mathtt{0}} > |u|_{\mathtt{3}}$ by assumption and the letters $\mathtt{0}$ and $\mathtt{3}$ alternate in $\mathbf{y}$ by Lemma 2.18. The conclusion is immediate. □

Lemma 2.35.

If n is odd and n_{12} is even, then $\mathcal{X}$ leads to only one 2-abelian equivalence class of $\mathbf{t}$ if and only if $n_{03} = \min_{03}(n)$ and $\mathrm{mj}_{03}(n) = 1$, or $n_{03} = \max_{03}(n)$ and $\mathrm{MJ}_{03}(n) = 1$. Otherwise, $\mathcal{X}$ splits into two classes.

Proof. If n is odd and n_{12} is even, then n_{03} is even. Assume that $n_{\mathtt{0}} > n_{\mathtt{3}}$ (the other case is symmetric). If $n_{03} = \min_{03}(n)$ and $\mathrm{mj}_{03}(n) = 1$ then, by Lemma 2.34, all the factors in $\mathcal{X}$ start with $\mathtt{0}$, and so $\mathcal{X}$ leads to only one class. If $n_{03} = \max_{03}(n)$ and $\mathrm{MJ}_{03}(n) = 1$, then all the factors in $\mathcal{X}$ are preceded and followed by $\mathtt{3}$. In particular, they all start with $\mathtt{2}$ and again $\mathcal{X}$ leads to only one class.

For the other direction, suppose that $\mathcal{X}$ leads to only one class. All the factors in $\mathcal{X}$ must start either with a letter in $\{\mathtt{0}, \mathtt{1}\}$ or with a letter in $\{\mathtt{2}, \mathtt{3}\}$. Assume first that all the elements of $\mathcal{X}$ start with $\mathtt{0}$ or $\mathtt{1}$. Let u be a factor in $\mathcal{X}$. If the first letter of u is $\mathtt{1}$, it must start with $\mathtt{120}$ since u has more $\mathtt{0}$'s than $\mathtt{3}$'s. Thus u is always preceded by $\mathtt{2}$. It cannot end with $\mathtt{1}$ (since $n_{\mathtt{1}} = n_{\mathtt{2}}$). So it must end with $\mathtt{0}$ or $\mathtt{2}$. If $u = \mathtt{120}u'\mathtt{2}$, then $\mathtt{2120}u'$ is an element of $\mathcal{X}$ starting with $\mathtt{2}$, which is a contradiction. If $u = \mathtt{120}u'\mathtt{0}$ then $u\mathtt{1}$ is a factor of $\mathbf{y}$. So $\mathtt{20}u'\mathtt{01}$ is an element of $\mathcal{X}$ starting with $\mathtt{2}$, a contradiction. Hence u cannot start with $\mathtt{1}$ and thus starts with $\mathtt{0}$. Observe that, if u does not end with $\mathtt{0}$, then $\tau(u)^{\mathrm{R}}$ is still an element of $\mathcal{X}$ by Lemma 2.19 and $\tau(u)^{\mathrm{R}}$ does not start with $\mathtt{0}$, a contradiction. Hence all the factors in $\mathcal{X}$ start and end with $\mathtt{0}$. By Lemma 2.34, we have $n_{03} = \max_{03}(n)$ and $\mathrm{MJ}_{03}(n) = 1$.

Assume now that all the elements of $\mathcal{X}$ start with $\mathtt{2}$ or $\mathtt{3}$. Since $n_{\mathtt{0}} > n_{\mathtt{3}}$, they all start with $\mathtt{2}$. Moreover, as $n_{\mathtt{1}} = n_{\mathtt{2}}$, they must end with $\mathtt{0}$ or $\mathtt{1}$. If $u \in \mathcal{X}$ ends with $\mathtt{0}$, then $\tau'(u)^{\mathrm{R}} \in \mathcal{X}$ starts with $\mathtt{3}$ by Lemma 2.19, a contradiction. So all factors in $\mathcal{X}$ end with $\mathtt{1}$. Let $u = \mathtt{2}u'\mathtt{1}$ be an element of $\mathcal{X}$. By Lemma 2.17, the only possible extensions of u as a factor of length $n+1$ of $\mathbf{y}$ are $\mathtt{1}u$, $\mathtt{3}u$, $u\mathtt{2}$ and $u\mathtt{3}$. If $\mathtt{1}u$ is a factor of $\mathbf{y}$, then $\mathtt{12}u' \in \mathcal{X}$ starts with $\mathtt{1}$, which is a contradiction. If $u\mathtt{2}$ is factor of $\mathbf{y}$, then $\tau(u'\mathtt{12})^{\mathrm{R}} \in \mathcal{X}$ starts with $\mathtt{1}$, a contradiction. Hence all the factors in $\mathcal{X}$ are preceded and followed by $\mathtt{3}$ in $\mathbf{y}$. By Lemma 2.34, it means that $n_{03} = \min_{03}(n)$ and $\mathrm{mj}_{03}(n) = 1$. □

We are now ready to prove Theorem 2.30.

Proof of Theorem 2.30. The difference between $\mathcal{P}^{(2)}_{\mathbf{t}}(n+1)$ and $\mathcal{P}^{(1)}_{\mathbf{y}}(n)$ is the number of abelian equivalence classes of factors of length n of $\mathbf{y}$ that split into two 2-abelian equivalence classes of factors of length $n+1$ of $\mathbf{t}$.

For even n, by Lemmas 2.32 and 2.33, it happens when n_{12} is even. The number of even values of $n_{12} \in \{\min_{12}(n), \dots, \max_{12}(n)\}$ is

$$\begin{cases} \frac{1}{2}\Delta_{12}(n) + 1 & \text{if } \min_{12}(n) \text{ and } \Delta_{12}(n) \text{ are even} \\ \frac{1}{2}\Delta_{12}(n) & \text{if } \min_{12}(n)+1 \text{ and } \Delta_{12}(n) \text{ are even} \\ \frac{1}{2}\Delta_{12}(n) + \frac{1}{2} & \text{if } \Delta_{12}(n) \text{ is odd,} \end{cases}$$

which leads to the result.

For odd n, by Lemmas 2.32 and 2.35, it happens when n_{12} is even, except if we have $n_{03} = \min_{03}(n)$ and $\mathrm{mj}_{03}(n) = 1$, or if $n_{03} = \max_{03}(n)$ and $\mathrm{MJ}_{03}(n) = 1$. The number of such cases is

$$\begin{cases} \frac{\Delta_{12}(n)}{2} + 1 - \mathrm{MJ}_{03}(n) - \mathrm{mj}_{03}(n) & \text{if } \min_{12}(n) \text{ and } \Delta_{12}(n) \text{ are even} \\ \frac{\Delta_{12}(n)+1}{2} - \mathrm{MJ}_{03}(n) & \text{if } \min_{12}(n) \text{ and } \Delta_{12}(n)+1 \text{ are even} \\ \frac{\Delta_{12}(n)+1}{2} - \mathrm{mj}_{03}(n) & \text{if } \min_{12}(n) \text{ and } \Delta_{12}(n) \text{ are odd} \\ \frac{\Delta_{12}(n)}{2} & \text{if } \min_{12}(n)+1 \text{ and } \Delta_{12}(n) \text{ are even.} \end{cases}$$

Indeed, consider for example the case where $\min_{12}(n)$ and $\Delta_{12}(n)$ are even. First, there are $\frac{\Delta_{12}(n)}{2} + 1$ even values of n_{12}. Second, since $\min_{12}(n)$ is even and n is odd, we have $\max_{03}(n) = n - \min_{12}(n)$ odd. Since $\Delta_{12}(n)$ is even, $\max_{12}(n)$ is also even and $\min_{03}(n)$ is odd.

If n is such that $\mathrm{mj}_{03}(n) = 1$ (resp. $\mathrm{MJ}_{03}(n) = 1$) then the case $n_{03} = \min_{03}(n)$ and $\mathrm{mj}_{03}(n) = 1$ (resp. $n_{03} = \max_{03}(n)$ and $\mathrm{MJ}_{03}(n) = 1$) indeed happens. So we have to remove 1, i.e., $\mathrm{mj}_{03}(n)$ or $\mathrm{MJ}_{03}(n)$ for each case.

As another example, consider the case where $\min_{12}(n)$ and $\Delta_{12}(n)$ are odd. Then $\max_{03}(n)$ is even and $\min_{03}(n)$ is odd. There are $\frac{\Delta_{12}(n)+1}{2}$ even values of n_{12}. We cannot have $n_{03} = \max_{03}(n)$ (for parity reasons) and thus we never have $n_{03} = \max_{03}(n)$ and $\mathrm{MJ}_{03}(n) = 1$. But the case $n_{03} = \min_{03}(n)$ happens and thus we have to remove one case when $\mathrm{mj}_{03}(n) = 1$.

Finally, observe that to each pair (n, n_{12}), with n odd and n_{12} even, corresponds two abelian equivalence classes of $\mathbf{y}$ (see the proof of Proposition 2.16). Each of these classes splits into two 2-abelian equivalence classes. Hence multiplying by 2 the number of pairs (n, n_{12}), with n odd and n_{12} even, gives the result claimed for n odd. □

Corollary 2.36.

The sequence $\mathcal{P}^{(2)}_{\mathbf{t}}(n)_{n\geq 0}$ is 2-regular.

Proof. We can make use of Lemma 1.34. Thanks to Theorem 2.30, $\mathcal{P}^{(2)}_{\mathbf{t}}(n+1)$ can be expressed as a combination of $\mathcal{P}^{(1)}_{\mathbf{y}}(n)$, $\Delta_{12}(n)$, $\mathrm{MJ}_{03}(n)$, $\mathrm{mj}_{03}(n)$ using the predicates (n mod 2), ($\Delta_{12}(n)$ mod 2) and ($\min_{12}(n)$ mod 2). From Corollary 2.21, all these predicates are 2-automatic. It remains to show that these functions are 2-regular.

- The sequence $\mathcal{P}_{\mathbf{y}}^{(1)}(n)_{n\geq 0}$ is 2-regular since $\mathcal{P}_{\mathbf{y}}^{(1)}$ is piecewise-defined by Proposition 2.16 and all functions (respectively predicates) occurring in the definition are 2-regular (resp. 2-automatic) by Corollary 2.21.
- The sequence $\Delta_{12}(n)_{n\geq 0}$ is 2-regular by Corollary 2.21.
- The sequences $\mathrm{MJ}_{03}(n)_{n\geq 0}$ and $\mathrm{mj}_{03}(n)_{n\geq 0}$ are 2-regular. Indeed, $\mathrm{MJ}_{03}(n+1)$ and $\mathrm{mj}_{03}(n)$ are linear combinations of $\min_{12}(n)$, $\min_{12}(n+1)$, $\Delta_{12}(n)$ and $\Delta_{12}(n+1)$. Moreover $\mathrm{MJ}_{03}(n+1)$ and $\mathrm{mj}_{03}(n)$ can only take the values 0 and 1. So the relations can be expressed using $(\min_{12}(n) \bmod 2)_{n\geq 0}$ and $(\Delta_{12}(n) \bmod 2)_{n\geq 0}$. Since these two latter sequences are 2-regular, the sequences $(\min_{12}(n+1) \bmod 2)_{n\geq 0}$ and $(\Delta_{12}(n+1) \bmod 2)_{n\geq 0}$ are 2-regular by Lemma 1.33 and so are $\mathrm{MJ}_{03}(n+1)_{n\geq 0}$ and $\mathrm{mj}_{03}(n)_{n\geq 0}$ by Lemma 1.34. Thus, $\mathrm{MJ}_{03}(n)_{n\geq 0}$ is 2-regular by Lemma 1.33.

Therefore, Lemma 1.34 implies that the sequence $\mathcal{P}_{\mathbf{t}}^{(2)}(n+1)_{n\geq 0}$ is 2-regular. Then the sequence $\mathcal{P}_{\mathbf{t}}^{(2)}(n)_{n\geq 0}$ is 2-regular by Lemma 1.33. $\square$

From Theorem 2.30, we can deduce recurrence relations satisfied by the 2-abelian complexity of the Thue–Morse word. In particular, we obtain the same result as the one obtained by Greinecker [Gre]: the sequence $\mathcal{P}_{\mathbf{t}}^{(2)}(n)_{n\geq 0}$ is a concatenation of longer and longer palindromes.

Proposition 2.37.

let $\ell \geq 2$ and $0 \leq r < 2^\ell$. We have

$$\mathcal{P}_{\mathbf{t}}^{(2)}(2^\ell + r + 1) = \mathcal{P}_{\mathbf{t}}^{(2)}(2^{\ell+1} - r + 1)$$

Proof. Let us first name the different cases from Theorem 2.30. We set

- Case A: r odd, $\min_{12}(2^\ell + r)$ even, $\Delta_{12}(2^\ell + r)$ even
- Case B: r odd, $\min_{12}(2^\ell + r)$ even, $\Delta_{12}(2^\ell + r)$ odd
- Case C: r odd, $\min_{12}(2^\ell + r)$ odd, $\Delta_{12}(2^\ell + r)$ odd
- Case D: r odd, $\min_{12}(2^\ell + r)$ odd, $\Delta_{12}(2^\ell + r)$ even
- Case E: r even, $\min_{12}(2^\ell + r)$ even, $\Delta_{12}(2^\ell + r)$ even
- Case F: r even, $\min_{12}(2^\ell + r)$ odd, $\Delta_{12}(2^\ell + r)$ even
- Case G: r even, $\Delta_{12}(2^\ell + r)$ odd.

The cases $A', \ldots, G'$ are defined similarly by replacing $2^\ell + r$ with $2^{\ell+1} - r$. We notice the following equivalences using Proposition 2.20:

- Case $A \Leftrightarrow$ Case A'
- Case $B \Leftrightarrow$ Case C'
- Case $C \Leftrightarrow$ Case B'
- Case $D \Leftrightarrow$ Case D'
- Case $E \Leftrightarrow$ Case E'
- Case $F \Leftrightarrow$ Case F'
- Case $G \Leftrightarrow$ Case G'.

We know that $\Delta_{12}(2^\ell + r) = \Delta_{12}(2^{\ell+1} - r)$ by Proposition 2.20 and that $\mathcal{P}^{(1)}_{\mathbf{y}}(2^\ell + r) = \mathcal{P}^{(1)}_{\mathbf{y}}(2^{\ell+1} - r)$ by Theorem 2.28. The last relation that we need is

$$\mathrm{mj}_{03}(2^\ell + r) = \mathrm{MJ}_{03}(2^{\ell+1} - r) \text{ and } \mathrm{MJ}_{03}(2^\ell + r) = \mathrm{mj}_{03}(2^{\ell+1} - r).$$

and it follows from Lemma 2.27. Now straightforward computation shows that

$$\mathcal{P}^{(2)}_{\mathbf{t}}(2^\ell + r + 1) = \mathcal{P}^{(2)}_{\mathbf{t}}(2^{\ell+1} - r + 1).$$

For instance, if Case B holds, Case C' holds too and we have

$$\begin{aligned}\mathcal{P}^{(2)}_{\mathbf{t}}(2^\ell + r + 1) &= \Delta_{12}(2^\ell + r) + 1 - 2\,\mathrm{MJ}(2^\ell + r)\\ &= \Delta_{12}(2^{\ell+1} - r) + 1 - 2\,\mathrm{mj}(2^{\ell+1} - r)\end{aligned}$$

which is equal to $\mathcal{P}^{(2)}_{\mathbf{t}}(2^{\ell+1} - r + 1) = \Delta_{12}(2^{\ell+1} - r) + 1 - 2\,\mathrm{mj}(2^{\ell+1} - r)$ □

2.4 The case of the period-doubling word p

In this section, we turn our attention to the period-doubling word $\mathbf{p}$. Note that $g(\mathbf{y})$ is exactly the period-doubling word $\mathbf{p}$, where g is the coding defined by $g(\mathtt{0}) = \mathtt{1}$, $g(\mathtt{1}) = \mathtt{0}$, $g(\mathtt{2}) = \mathtt{0}$ and $g(\mathtt{3}) = \mathtt{1}$. In particular, $\Delta_{12}(n) + 1$ is the abelian complexity function of the period-doubling word. This function was also studied in [BSCRF14, KSZ]. Here we obtain relations of the same type as the relations in Theorem 2.11.

We let $\mathbf{x}$ denote

$$\mathrm{block}(\mathbf{p}, 2) = \mathtt{120012121200120012001212120012\cdots}\,,$$

the 2-block coding of $\mathbf{p}$, introduced in Example 1.49. Recall that $\mathbf{x}$ is the fixed point of the morphism η defined by $\eta : \mathtt{0} \mapsto \mathtt{12}, \mathtt{1} \mapsto \mathtt{12}, \mathtt{2} \mapsto \mathtt{00}$. The approach here is similar to that of the Thue–Morse word: we consider in this section the abelian complexity $\mathcal{P}^{(1)}_{\mathbf{x}}(n)$ of $\mathbf{x}$ and then we compare $\mathcal{P}^{(1)}_{\mathbf{x}}(n)$ with the 2-abelian complexity $\mathcal{P}^{(2)}_{\mathbf{p}}(n)$ of $\mathbf{p}$.

Our study of the Thue–Morse word in Section 2.3 made substantial use of counting $\mathtt{1}$'s and $\mathtt{2}$'s in factors of $\mathbf{y}$. Alternatively, we could have counted the total number of $\mathtt{0}$'s and $\mathtt{3}$'s in factors of $\mathbf{y}$, since this is equivalent information and since the letters 0 and 3 alternate in $\mathbf{y}$.

For the period-doubling word, the appropriate statistic for factors of $\mathbf{x}$ is the total number of $\mathtt{0}$'s (or, equivalently, the total number of $\mathtt{1}$'s and $\mathtt{2}$'s). Let $n \in \mathbb{N}$. We let $\max_0(n)$ (resp. $\min_0(n)$) denote the maximum (resp. minimum) number of $\mathtt{0}$'s in a factor of $\mathbf{x}$ of length n. Let $\Delta_0(n) = \max_0(n) - \min_0(n)$ be the difference between these two values.

Each of the $\Delta_0(n)+1$ integers in the interval $[\min_0(n), \max_0(n)]$ is attained as the number of $\mathtt{0}$'s in some factor of $\mathbf{x}$ of length n. Indeed, when we slide a window of length n along $\mathbf{x}$ from a factor with $\min_0(n)$ zeros to a factor with $\max_0(n)$ zeros, the number of $\mathtt{0}$'s changes by at most 1 per step.

Lemma 2.38.

If n is even, then $\max_0(n)$, $\min_0(n)$ and $\Delta_0(n)$ are even.

Proof. Suppose a factor $w = w_1 \cdots w_{2n}$ of $\mathbf{x}$ of even length $2n$ has an odd number n_0 of zeros. Since $\eta(\mathtt{0}) = \eta(\mathtt{1}) = \mathtt{12}$ and $\eta(\mathtt{2}) = \mathtt{00}$, the factor w starts or ends with $\mathtt{0}$. Without loss of generality, assume it starts with $w_1 = \mathtt{0}$. Then its last letter must be $w_{2n} = \mathtt{1}$. The words $\mathtt{0}w_1 \cdots w_{2n-1}$ and $w_2 \cdots w_{2n}\mathtt{2}$ are two factors of length $2n$ with respectively $n_0 + 1$ and $n_0 - 1$ zeros. Hence, these two factors have even numbers of zeros which are respectively greater than and less than n_0. The conclusion follows. ☐

We give two related proofs of the 2-regularity of the sequence $\mathcal{P}^{(1)}_{\mathbf{x}}(n)_{n\geq 0}$. The first uses the fact that $\mathcal{P}^{(1)}_{\mathbf{x}}(n)_{n\geq 0}$ is piecewise-defined, together with the fact that the function occurring in this definition, namely $\Delta_0(n)_{n\geq 0}$, is 2-regular and the two sequences $(\Delta_0(n) \bmod 2)_{n\geq 0}$ and $(\min_0(n) \bmod 2)_{n\geq 0}$ are 2-automatic. Then the 2-regularity of the sequence $\mathcal{P}^{(1)}_{\mathbf{x}}(n)_{n\geq 0}$ will follow from Lemma 1.34. In the second proof, we show that the abelian complexity of $\mathbf{x}$ satisfies a reflection symmetry, which allows us to apply our general result expressed by Theorem 2.11. Finally, we show that the 2-regularity of $\mathcal{P}^{(2)}_{\mathbf{t}}(n)_{n\geq 0}$ follows from the 2-regularity of $\Delta_0(n)_{n\geq 0}$ and $\mathcal{P}^{(1)}_{\mathbf{x}}(n)_{n\geq 0}$.

2.4.1 The abelian complexity of $\mathrm{block}(\mathbf{p}, 2)$ is 2-regular

Following the same approach as in Subsection 2.3.1, one can show that the abelian complexity of the 2-block coding of $\mathbf{p}$ is piecewise-defined. Details of the proof are given in Appendix A.

Proposition 2.39.

For $n \in \mathbb{N}$,

$$\mathcal{P}^{(1)}_{\mathbf{x}}(n) = \begin{cases} \frac{3}{2}\Delta_0(n) + \frac{3}{2} & \text{if } \Delta_0(n) \text{ is odd} \\ \frac{3}{2}\Delta_0(n) + 1 & \text{if } \Delta_0(n) \text{ and } n - \min_0(n) \text{ are even} \\ \frac{3}{2}\Delta_0(n) + 2 & \text{if } \Delta_0(n) \text{ and } n - \min_0(n) + 1 \text{ are even.} \end{cases}$$

Moreover the sequences occurring in this piecewise definition satisfy recurrence relations similar to the assumptions of Theorem 2.11. A complete proof of the following proposition is available in Appendix A.

Proposition 2.40.

Let $\ell \geq 2$ and $0 \leq r < 2^\ell$. We have

$$\Delta_0(2^\ell + r) = \begin{cases} \Delta_0(r) + 2 & \text{if } r \leq 2^{\ell-1} \\ \Delta_0(2^{\ell+1} - r) & \text{if } r > 2^{\ell-1} \end{cases}$$

and

$$\min_0(2^\ell + r) \equiv \begin{cases} \min_0(r) \pmod 2 & \text{if } r \leq 2^{\ell-1} \\ \min_0(2^{\ell+1} - r) + \Delta_0(2^{\ell+1} - r) \pmod 2 & \text{if } r > 2^{\ell-1}. \end{cases}$$

Then the 2-regularity of $\mathcal{P}_{\mathbf{x}}^{(1)}(n)$ follows from Lemma 1.34 since all sequences (respectively predicates) occurring in the definition are 2-regular (resp. 2-automatic). Recall that a k-regular sequence taking only finitely many values is k-automatic.

Corollary 2.41.

- The sequence $\Delta_0(n)_{n\geq 0}$ is 2-regular.
- The sequence $(\Delta_0(n) \bmod 2)_{n\geq 0}$ is 2-automatic.
- The sequence $(\min_0(n) \bmod 2)_{n\geq 0}$ is 2-automatic.

Proof. The first assertion is a direct consequence of Theorem 2.11 and Proposition 2.40. Note that one can obtain explicit relations satisfied by $\Delta_0(n)_{n\geq 0}$ from Example 2.14. The second assertion follows from Lemma 1.31.

For the last assertion, for $i \in \{0, \dots, 31\}$ we prove that, modulo 2,

$$\min_0(32n+i) \equiv \begin{cases} \min_0(8n+1) & \text{if } i \in \{1,5,9,17,25\} \\ \min_0(8n+3) & \text{if } i = 11 \\ \min_0(8n+5) & \text{if } i = 21 \\ \min_0(8n+7) & \text{if } i \in \{7,15,23,27,31\} \\ 0 & \text{otherwise} \end{cases}$$

and

$$\Delta_0(32n+i) \equiv \begin{cases} \Delta_0(8n+1) & \text{if } i \in \{1,5,9,17,25\} \\ \Delta_0(8n+3) & \text{if } i = 11 \\ \Delta_0(8n+5) & \text{if } i = 21 \\ \Delta_0(8n+7) & \text{if } i \in \{7,15,23,27,31\} \\ 0 & \text{otherwise.} \end{cases}$$

By Lemma 2.38, we already know that $\min_0(2n) \equiv \Delta_0(2n) \equiv 0 \pmod 2$ for any $n \in \mathbb{N}$. Hence the relations above are true for i even. We prove the other relations by induction on n. They are true for $n = 0$. Let $n > 0$ and assume the relations are satisfied for all n' such that $0 \leq n' < n$. We can write $n = 2^\ell + r$ with $\ell \geq 0$ and $0 \leq r < 2^\ell$. Let $i \in \{1, \dots, 31\}$ be odd.

Assume first that $r < 2^{\ell-1}$. We have $32n + i = 2^{\ell+5} + 32r + i$ and $32r + i < 2^{\ell+4}$.

$$\begin{aligned} \min_0(32n+i) &\equiv \min_0(32r+i) && \text{(by Proposition 2.40)} \\ &\equiv \min_0(8r+j) && \text{(by inductive hypothesis)} \\ &\equiv \min_0(2^{\ell+3}+8r+j) && \text{(by Proposition 2.40)} \\ &\equiv \min_0(8n+j) \pmod 2 \end{aligned}$$

for some $j \in \{0, \dots, 7\}$ according to the relations. A similar reasoning holds for the Δ_0 relations.

Assume now that $r \geq 2^{\ell-1}$. Since $32r + i > 2^{\ell+4}$, we have

$$\begin{aligned} \min_0(32n+i) &\equiv \min_0(2^{\ell+6}-32r-i) + \Delta_0(2^{\ell+6}-32r-i) && \text{(by Proposition 2.40)} \\ &\equiv \min_0(32n'+j) + \Delta_0(32n'+j) \pmod 2 \end{aligned}$$

with $j = 32 - i$ and $n' = 2^{\ell+1} - r - 1$. If $i \in \{3, 13, 19, 29\}$, then $j \in \{3, 13, 19, 29\}$. By the inductive hypothesis, $\min_0(32n' + j) \equiv \Delta_0(32n' + j) \equiv 0 \pmod 2$ and we are done.

For the remaining cases, $i, j \notin \{3, 13, 19, 29\}$. As $\min_0$ and Δ_0 satisfy the same recurrence relations, by the inductive hypothesis, there exists $k \in \{1, 3, 5, 7\}$ such that

$$\begin{aligned}\min_0(32n + i) &\equiv \min_0(8n' + k) + \Delta_0(8n' + k) \\ &\equiv \min_0(2^{\ell+4} - (8r + 8 - k)) + \Delta_0(2^{\ell+4} - (8r + 8 - k)) \\ &\equiv \min_0(2^{\ell+3} + (8r + 8 - k)) && \text{(by Proposition 2.40)} \\ &\equiv \min_0(8n + (8 - k)) \pmod 2.\end{aligned}$$

Observe that the value of $8 - k$ is the value given in the relation for i. This concludes the proof of the $\min_0$ relations. A similar argument works for the Δ_0 relations. □

2.4.2 The abelian complexity of $\mathrm{block}(\mathbf{p}, 2)$ satisfies a reflection symmetry and so is 2-regular

In this subsection we prove the 2-regularity of the abelian complexity $\mathcal{P}^{(1)}_{\mathbf{x}}(n)_{n\geq 0}$ in a second way. As we did for the proof of Theorem 2.28, we can express recurrence relations for $\mathcal{P}^{(1)}_{\mathbf{x}}$ using Propositions 2.39 and 2.40. In this case, these recurrence relations coincide with the framework of Theorem 2.11.

Theorem 2.42.

Let $\ell \geq 2$ and $0 \leq r < 2^\ell$. We have

$$\mathcal{P}^{(1)}_{\mathbf{x}}(2^\ell + r) = \begin{cases} \mathcal{P}^{(1)}_{\mathbf{x}}(r) + 3 & \text{if } r \leq 2^{\ell-1} \\ \mathcal{P}^{(1)}_{\mathbf{x}}(2^{\ell+1} - r) & \text{if } r > 2^{\ell-1}. \end{cases}$$

In particular, the sequence $\mathcal{P}^{(1)}_{\mathbf{x}}(n)_{n\geq 0}$ is 2-regular.

From Theorem 2.42, we see that $\mathcal{P}^{(1)}_{\mathbf{x}}(2^\ell) = \mathcal{P}^{(1)}_{\mathbf{x}}(0) + 3 = 4$ for all $\ell \geq 2$. Additionally, one can check that $\mathcal{P}^{(1)}_{\mathbf{x}}(2^1) = 4$.

Proof of Theorem 2.42. If $2^{\ell-1} \leq r \leq 2^\ell$, since all the conditions in Proposition 2.39 are equivalent whether considering $2^\ell + r$ or $2^{\ell+1} - r$, we have

$$\mathcal{P}^{(1)}_{\mathbf{x}}(2^\ell + r) = \mathcal{P}^{(1)}_{\mathbf{x}}(2^{\ell+1} - r).$$

Assume now that $0 \leq r \leq 2^{\ell-1}$. If $\Delta_0(2^\ell + r)$ is odd, $\Delta_0(r)$ is also odd by Proposition 2.40. By Proposition 2.39, we have $\mathcal{P}^{(1)}_{\mathbf{x}}(2^\ell + r) = \frac{3}{2}(\Delta_0(2^\ell + r) + 1)$ and $\mathcal{P}^{(1)}_{\mathbf{x}}(r) = \frac{3}{2}(\Delta_0(r) + 1)$. By Proposition 2.40, we have $\Delta_0(2^\ell + r) = \Delta_0(r) + 2$. Putting these three equalities together, we get $\mathcal{P}^{(1)}_{\mathbf{x}}(2^\ell + r) = \mathcal{P}^{(1)}_{\mathbf{x}}(r) + 3$.

The other cases can be done similarly. If $\Delta_0(2^\ell + r)$ and $2^\ell + r - \min_0(2^\ell + r)$ are even, then $\Delta_0(r)$ and $r - \min_0(r)$ are even and

$$\begin{aligned}\mathcal{P}^{(1)}_{\mathbf{x}}(2^\ell + r) &= \tfrac{3}{2}\Delta_0(2^\ell + r) + 1 && \text{(by Proposition 2.39)} \\ &= \tfrac{3}{2}(\Delta_0(r) + 2) + 1 && \text{(by Proposition 2.40)} \\ &= \mathcal{P}^{(1)}_{\mathbf{x}}(r) + 3 && \text{(by Proposition 2.39).}\end{aligned}$$

If $\Delta_0(2^\ell+r)$ is even and $2^\ell+r-\min_0(2^\ell+r)$ is odd, then $\Delta_0(r)$ is even and $r-\min_0(r)$ is odd. Then

$$\begin{aligned}\mathcal{P}^{(1)}_{\mathbf{x}}(2^\ell+r) &= \tfrac{3}{2}\Delta_0(2^\ell+r)+2 && \text{(by Proposition 2.39)}\\ &= \tfrac{3}{2}(\Delta_0(r)+2)+2 && \text{(by Proposition 2.40)}\\ &= \mathcal{P}^{(1)}_{\mathbf{x}}(r)+3 && \text{(by Proposition 2.39).}\end{aligned}$$ □

One can prove the following result in a manner similar to the proof of Theorem 2.11. There may be simpler recurrences, but these relations exhibit the same symmetry as in Theorem 2.11.

Theorem 2.43.

The abelian complexity sequence $\mathcal{P}^{(1)}_{\mathbf{x}}(n)_{n\geq 0}$ of the 2-block coding of the period-doubling word satisfies the following relations.

$$\begin{aligned}
\mathcal{P}^{(1)}_{\mathbf{x}}(8n) &= \mathcal{P}^{(1)}_{\mathbf{x}}(2n)\\
4\mathcal{P}^{(1)}_{\mathbf{x}}(8n+1) &= -2\mathcal{P}^{(1)}_{\mathbf{x}}(2n+1)+7\mathcal{P}^{(1)}_{\mathbf{x}}(4n+1)-2\mathcal{P}^{(1)}_{\mathbf{x}}(4n+2)+\mathcal{P}^{(1)}_{\mathbf{x}}(4n+3)\\
4\mathcal{P}^{(1)}_{\mathbf{x}}(8n+2) &= -6\mathcal{P}^{(1)}_{\mathbf{x}}(2n+1)+9\mathcal{P}^{(1)}_{\mathbf{x}}(4n+1)-2\mathcal{P}^{(1)}_{\mathbf{x}}(4n+2)+3\mathcal{P}^{(1)}_{\mathbf{x}}(4n+3)\\
4\mathcal{P}^{(1)}_{\mathbf{x}}(8n+3) &= -6\mathcal{P}^{(1)}_{\mathbf{x}}(2n+1)+5\mathcal{P}^{(1)}_{\mathbf{x}}(4n+1)+2\mathcal{P}^{(1)}_{\mathbf{x}}(4n+2)+3\mathcal{P}^{(1)}_{\mathbf{x}}(4n+3)\\
\mathcal{P}^{(1)}_{\mathbf{x}}(8n+4) &= \mathcal{P}^{(1)}_{\mathbf{x}}(4n+2)\\
4\mathcal{P}^{(1)}_{\mathbf{x}}(8n+5) &= -6\mathcal{P}^{(1)}_{\mathbf{x}}(2n+1)+3\mathcal{P}^{(1)}_{\mathbf{x}}(4n+1)+2\mathcal{P}^{(1)}_{\mathbf{x}}(4n+2)+5\mathcal{P}^{(1)}_{\mathbf{x}}(4n+3)\\
4\mathcal{P}^{(1)}_{\mathbf{x}}(8n+6) &= -6\mathcal{P}^{(1)}_{\mathbf{x}}(2n+1)+3\mathcal{P}^{(1)}_{\mathbf{x}}(4n+1)-2\mathcal{P}^{(1)}_{\mathbf{x}}(4n+2)+9\mathcal{P}^{(1)}_{\mathbf{x}}(4n+3)\\
4\mathcal{P}^{(1)}_{\mathbf{x}}(8n+7) &= -2\mathcal{P}^{(1)}_{\mathbf{x}}(2n+1)+\mathcal{P}^{(1)}_{\mathbf{x}}(4n+1)-2\mathcal{P}^{(1)}_{\mathbf{x}}(4n+2)+7\mathcal{P}^{(1)}_{\mathbf{x}}(4n+3)
\end{aligned}$$

2.4.3 The 2-abelian complexity of p is 2-regular

As for the Thue–Morse word, we express the 2-abelian complexity $\mathcal{P}^{(2)}_{\mathbf{p}}$ in terms of the 1-abelian complexity $\mathcal{P}^{(1)}_{\mathbf{x}}$, $\Delta_0(n)$ and two new functions $\mathrm{MJ}_0(n)$ and $\mathrm{mj}_0(n)$, in order to prove the 2-regularity of $\mathcal{P}^{(2)}_{\mathbf{p}}(n)_{n\geq 0}$.

The functions $\mathrm{MJ}_0 : \mathbb{N} \to \{0,1\}$ and $\mathrm{mj}_0 : \mathbb{N} \to \{0,1\}$ are defined analogously to the functions $\mathrm{MJ}_{03}(n)$ and $\mathrm{mj}_{03}(n)$ of Subsection 2.3.2. We set $\mathrm{MJ}_0(0)=0$ and, for $n\geq 1$,

$$\mathrm{MJ}_0(n) := \begin{cases} 1 & \text{if } \max_0(n) > \max_0(n-1)\\ 0 & \text{otherwise,}\end{cases}$$

Similarly, for $n \geq 0$,

$$\mathrm{mj}_0(n) := \begin{cases} 1 & \text{if } \min_0(n+1) > \min_0(n)\\ 0 & \text{otherwise.}\end{cases}$$

Since $\max_0(n)$ and $\min_0(n)$ are non-decreasing, we can write

$$\begin{aligned}\mathrm{MJ}_0(n+1) &= \max_0(n+1)-\max_0(n),\\ \mathrm{mj}_0(n) &= \min_0(n+1)-\min_0(n).\end{aligned}$$

To compute $\mathcal{P}_{\mathbf{p}}^{(2)}$, we use the abelian complexity of $\mathbf{x} = \mathrm{block}(\mathbf{p}, 2)$, $\mathcal{P}_{\mathbf{x}}^{(1)}$, and study when an abelian equivalence class of length-n factors of $\mathbf{x}$ splits into one or two 2-abelian equivalence classes of factors of length $n+1$ of $\mathbf{p}$. The relationship between these sequences and $\mathcal{P}_{\mathbf{p}}^{(2)}$ and $\mathcal{P}_{\mathbf{x}}^{(1)}$ is stated in the following result.

Proposition 2.44.

Let $n \geq 1$ be an integer. Then

$$\mathcal{P}_{\mathbf{p}}^{(2)}(n+1) - \mathcal{P}_{\mathbf{x}}^{(1)}(n) = \begin{cases} 0 & \text{if } n \text{ is odd} \\ \frac{\Delta_0(n)}{2} + 1 - \mathrm{MJ}_0(n) - \mathrm{mj}_0(n) & \text{if } n \text{ is even.} \end{cases}$$

The proof of Proposition 2.44 follows the same ideas used in Subsection 2.3.2 and is available in Appendix A.

Corollary 2.45.

The sequence $\mathcal{P}_{\mathbf{p}}^{(2)}(n)_{n\geq 0}$ is 2-regular.

Proof. We can make use of Lemma 1.34. Thanks to Proposition 2.44, $\mathcal{P}_{\mathbf{p}}^{(2)}(n+1)$ can be expressed as a combination of $\mathcal{P}_{\mathbf{x}}^{(1)}(n)$, $\Delta_0(n)$, $\mathrm{MJ}_0(n)$, $\mathrm{mj}_0(n)$ using the predicate $(n \bmod 2)$. Note that the predicate $(n \bmod 2)$ is trivially 2-automatic.

- The sequence $\mathcal{P}_{\mathbf{x}}^{(1)}(n)_{n\geq 0}$ is 2-regular by Theorem 2.42.
- The sequence $\Delta_0(n)_{n\geq 0}$ is 2-regular by Corollary 2.41.
- Both sequences $\mathrm{MJ}_0(n)_{n\geq 0}$ and $\mathrm{mj}_0(n)_{n\geq 0}$ are 2-regular. Indeed, observe that

$$\begin{aligned} \mathrm{MJ}_0(n+1) &= \max\nolimits_0(n+1) - \max\nolimits_0(n) \\ &= \min\nolimits_0(n+1) + \Delta_0(n+1) - \min\nolimits_0(n) - \Delta_0(n). \end{aligned}$$

  Since $\mathrm{MJ}_0(n+1)$ can only take the values 0 and 1, the latter relation can also be expressed using $(\min_0(n) \bmod 2)_{n\geq 0}$ and $(\Delta_0(n) \bmod 2)_{n\geq 0}$. These latter sequences are 2-regular by Corollary 2.41. By Lemma 1.33, $\mathrm{MJ}_0(n+1)_{n\geq 0}$ is thus a combination of four 2-regular sequences. Applying again Lemma 1.33, $\mathrm{MJ}_0(n)_{n\geq 0}$ is also 2-regular. We can show similarly that $\mathrm{mj}_0(n)_{n\geq 0}$ is 2-regular. In fact, both sequences $\mathrm{MJ}_0(n)_{n\geq 0}$ and $\mathrm{mj}_0(n)_{n\geq 0}$ are 2-automatic since they only take values 0 and 1.

Thus, all the functions in the expression for $\mathcal{P}_{\mathbf{p}}^{(2)}(n+1)$ are 2-regular. Finally, as the sequence $\mathcal{P}_{\mathbf{p}}^{(2)}(n+1)_{n\geq 0}$ is 2-regular, the sequence $\mathcal{P}_{\mathbf{p}}^{(2)}(n)_{n\geq 0}$ is 2-regular by Lemma 1.33. □

2.5 Conclusions and perspectives

The two examples treated in this chapter, namely the 2-abelian complexity of the Thue–Morse word and the period-doubling word, suggest that a general framework to study the

ℓ-abelian complexity of k-automatic sequences may exist. As an example, we consider the 3-block coding of the period-doubling word $\mathbf{p}$,

$$\mathbf{z} = \text{block}(\mathbf{p}, 3) = \mathtt{240125252401240124} \cdots .$$

The abelian complexity $\mathcal{P}^{(1)}_{\mathbf{z}}(n)_{n\geq 0} = (1, 5, 5, 8, 6, 10, 19, 11, \ldots)$ (Figure 2.12) seems to satisfy, for $\ell \geq 4$, the following relations (which are quite similar to what we have discussed so far)

$$\mathcal{P}^{(1)}_{\mathbf{z}}(2^\ell + r) = \begin{cases} \mathcal{P}^{(1)}_{\mathbf{z}}(r) + 5 & \text{if } r \leq 2^{\ell-1} \text{ and } r \text{ even} \\ \mathcal{P}^{(1)}_{\mathbf{z}}(r) + 7 & \text{if } r \leq 2^{\ell-1} \text{ and } r \text{ odd} \\ \mathcal{P}^{(1)}_{\mathbf{z}}(2^{\ell+1} - r) & \text{if } r > 2^{\ell-1}. \end{cases}$$

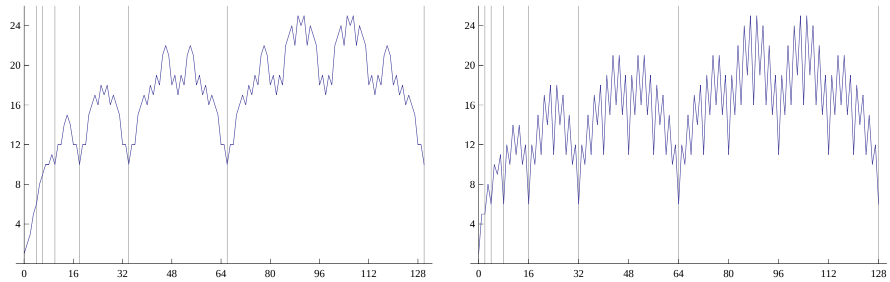


Figure 2.12: The 3-abelian complexity of the period-doubling word $\mathbf{p}$ (on the left) and the abelian complexity of block$(\mathbf{p}, 3)$ (on the right).

Then, the next step would be to relate $\mathcal{P}^{(3)}_{\mathbf{p}}(n+2)$ with $\mathcal{P}^{(1)}_{\mathbf{z}}(n)$ (and try to extend the developments from Section 2.3 and Section 2.4). We conjecture that the 3-abelian complexity of the period-doubling word satisfies the following relations

$$\mathcal{P}^{(3)}_{\mathbf{p}}(2^\ell + r + 2) = \begin{cases} \mathcal{P}^{(3)}_{\mathbf{p}}(r+2) + 6 & \text{if } r \leq 2^{\ell-1} \text{ and } r \equiv 2 \pmod 4 \\ \mathcal{P}^{(3)}_{\mathbf{p}}(r+2) + 7 & \text{if } r \leq 2^{\ell-1} \text{ and } r \not\equiv 2 \pmod 4 \\ \mathcal{P}^{(3)}_{\mathbf{p}}(2^{\ell+1} - r + 2) & \text{if } r > 2^{\ell-1}. \end{cases}$$

Hence, it satisfies a reflection symmetry over intervals of the form $[2^\ell + 2, 2^{\ell+1} + 2]$.

For the Thue–Morse word $\mathbf{t}$, the 3-abelian complexity is depicted in Figure 2.13 along with the abelian complexity of the 3-block coding of $\mathbf{t}$. Note that the two graphs look alike but different scales were used for the y coordinates. It seems that both functions satisfy a reflection symmetry respectively over the intervals of the form $[2^\ell+2, 2^{\ell+1}+2]$ and $[2^\ell, 2^{\ell+1}]$. But we cannot easily guess which recurrence relations hold for these two functions.

Recently, this work was presented at *Mons Days 2014*. The talk raised many questions and comments. Since the abelian complexity functions we considered are oscillating, it could be interesting to study the behaviour of the average of the functions. Moreover, we could look at each interval over which the taken values satisfy a reflection symmetry, and then rescale all these "windows" in order to compare them. For instance, we could check whether they converge to a continuous non-differentiable function.

Note that both the Thue–Morse word and the period-doubling word are palindromic, i.e., the set of all factors is closed under reversal. One can think the reflection symmetry is

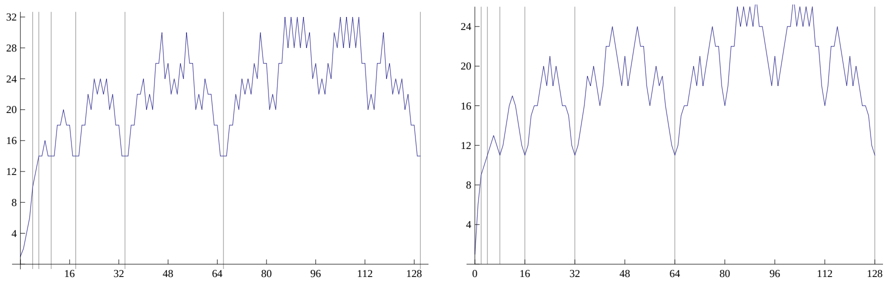
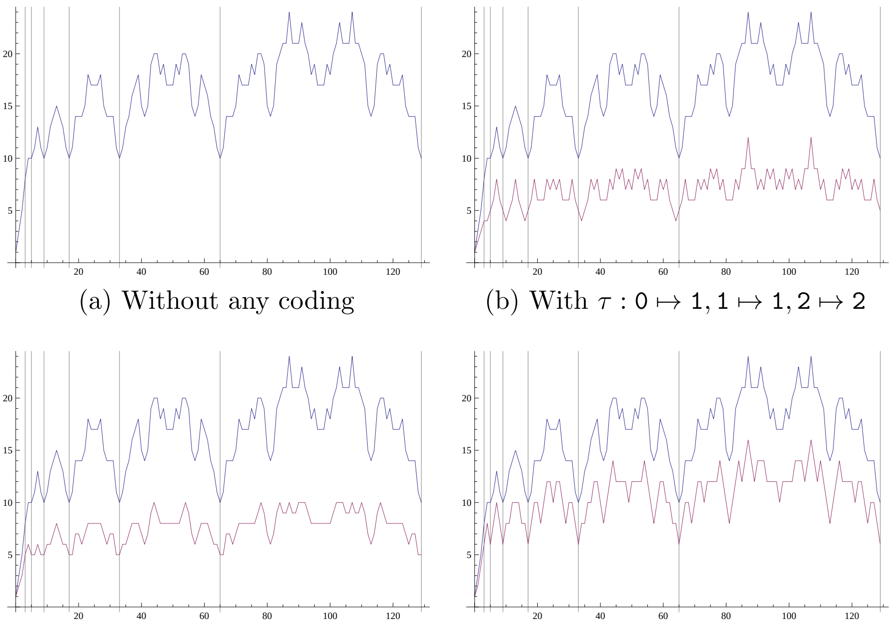


Figure 2.13: The 3-abelian complexity of the Thue–Morse word $\mathbf{t}$ (on the left) and the abelian complexity of block$(\mathbf{t}, 3)$ (on the right).

linked to palindromicity. It is not the case since the 2-block coding of the Thue–Morse word (respectively the period-doubling word) is not palindromic (for instance `01` is a factor but not 10) and its abelian complexity satisfies a reflection symmetry. Still both 2-block codings are closed under the application of reversal composed with the coding τ or τ' (Lemmas 2.19 and A.4). It would be interesting to see if the words presented in Appendix A, which have an abelian or a 2-abelian complexity satisfying a reflection symmetry, are closed for reversal composed with a coding.

Also we have only considered 2-automatic sequences that are pure morphic. Two possible extensions of our work are to consider morphic words $\tau(\varphi^\omega(a))$ with a non-uniform morphism φ, or 2-automatic sequences that are not pure morphic. In the last case, some computer experiments suggest that for a word $\mathbf{w}$ generated by a 2-uniform morphism, if the 2-abelian complexity of $\mathbf{w}$ satisfies a reflection symmetry, then so does the 2-abelian complexity of any coding of $\mathbf{w}$ (Figure 2.14).



(a) Without any coding

(b) With $\tau : \mathtt{0} \mapsto \mathtt{1}, \mathtt{1} \mapsto \mathtt{1}, \mathtt{2} \mapsto \mathtt{2}$

(c) With $\tau : \mathtt{0} \mapsto \mathtt{2}, \mathtt{1} \mapsto \mathtt{1}, \mathtt{2} \mapsto \mathtt{2}$

(d) With $\tau : \mathtt{0} \mapsto \mathtt{0}, \mathtt{1} \mapsto \mathtt{2}, \mathtt{2} \mapsto \mathtt{2}$

Figure 2.14: The 2-abelian complexity functions of the fixed point $\mathbf{w} \in \{\mathtt{0}, \mathtt{1}, \mathtt{2}\}^{\mathbb{N}}$ of the morphism defined by $\mathtt{0} \mapsto \mathtt{01}, \mathtt{1} \mapsto \mathtt{20}, \mathtt{2} \mapsto \mathtt{10}$ (upper curve) and of $\tau(\mathbf{w})$ where τ is a coding (lower curve).

Chapter 3

Abelian return words

We investigate some properties of abelian return words as recently introduced by Puzynina and Zamboni. In particular, we obtain a characterization of Sturmian words with non-zero intercept in terms of the finiteness of the set of abelian return words to all prefixes. We describe this set of abelian returns for the Fibonacci word but also for the 2-automatic Thue–Morse word. We also investigate the relationship existing between abelian complexity and finiteness of the set of abelian returns to all prefixes. The work presented in this chapter is based on a collaboration with my co-advisor Rigo and a postdoctoral fellow Salimov.

Contents

The return word is a classical notion in combinatorics on words and symbolic dynamical systems [Dur98, HZ99, JV00, Vui01]. For instance, Durand obtained a characterization of primitive substitutive sequences in terms of return words and derived sequences (Theorem 1.58). Vuillon obtained a characterization of Sturmian words in terms of return words (Theorem 1.59. Recall from Definition 1.55, that a return word to a factor u in an infinite word $\mathbf{x}$ is a factor of $\mathbf{x}$ that starts with u and ends before the next occurrence of u in $\mathbf{x}$.

In this chapter, we consider the abelian analogue of this notion of return word, which are called abelian return words or simply abelian returns. Such a study has been recently presented by Puzynina and Zamboni during the WORDS 2011 conference. Here we focus on different aspects of abelian returns and we hope that our results can be seen as complementary to those found by Puzynina and Zamboni [PZ13]. Their main contribution is a characterization of Sturmian words similar to the one obtained by Vuillon.

Theorem 3.1. [PZ13]

An aperiodic recurrent infinite word is Sturmian if and only if each of its factors has two or three abelian returns.

Puzynina and Zamboni also discuss the link between the number of abelian returns and periodicity. They provide a sufficient condition for periodicity.

Lemma 3.2. [PZ13]

Let k denote the size of the alphabet A. If each factor of a recurrent infinite word over the alphabet A has a most k abelian returns, then the word is periodic.

The main difference between [PZ13] and this present work is that we usually consider the set of abelian returns with respect to all the factors of an infinite word $\mathbf{x}$, while Puzynina and Zamboni study the set of abelian returns with respect to each factor taken separately.

This chapter, which is based on a joint work [RSV13] with my co-advisor Rigo and a post-doctoral fellow at University of Liège Salimov, is organized as follows. In Section 3.1, we present the main definitions and notation used in this chapter.

In Section 3.2, we discuss the relationship with periodicity and we prove that a recurrent word is periodic if and only if its set of abelian returns is finite. We also construct an abelian uniformly recurrent word which is not uniformly recurrent.

In Section 3.3, we restrict ourselves to the set $\mathcal{APR}_{\mathbf{x}}$ of abelian returns to all prefixes. In particular, this set is finite for any uniformly recurrent and abelian periodic word. We study the special case of the Thue–Morse word $\mathbf{t}$ introduced in Example 1.11. We show that the set of abelian returns to all prefixes of $\mathbf{t}$ contains 16 elements. Next, we obtain a characterization of Sturmian words with (non)zero intercept as follows. Let $\mathbf{x} = ST(\alpha, \rho)$ be a Sturmian word coding an orbit $(R_\alpha^n(\rho))_{n\geq 0}$. The set $\mathcal{APR}_{\mathbf{x}}$ of abelian returns to the prefixes of $\mathbf{x}$ is finite if and only if $\mathbf{x}$ does not have a null intercept (see Theorem 3.18). The celebrated Fibonacci word $\mathbf{f}$ introduced in Example 1.14 can be defined with a slope and an intercept both equal to $1/\phi^2$ where ϕ is the Golden mean. Therefore our result implies that $\mathcal{APR}_{\mathbf{f}}$ is finite. We show that this set contains exactly 5 elements. Interestingly, our developments can be related to the lengths of the palindromic prefixes of $\mathbf{f}$. See for instance [dL97, Fis06]. By contrast the set of abelian returns to all prefixes for the word $0\mathbf{f}$ is infinite. Then we show that if $\mathbf{x}$ is an abelian recurrent word such that $\mathcal{APR}_{\mathbf{x}}$ is finite, then $\mathbf{x}$ has bounded abelian complexity.

In Section 3.4, we introduce the notion of abelian derived sequence. If a word $\mathbf{x}$ is uniformly recurrent, then $\mathbf{x}$ can be factored in terms of abelian returns to a given prefix of $\mathbf{x}$. This gives rise to a coding that allows one to define a new sequence. Contrary to the non-abelian case and the characterization obtained by Durand (Theorem 1.58), the Thue–Morse word is an example of word having infinitely many abelian derived sequences.

In the last section of this chapter, we present related work that extends or completes our result.

3.1 Abelian return words

Recently, the notion of return words has been generalized to an abelian framework [PZ13]. We will distinguish two cases: abelian return to a prefix and abelian return to a factor. We make such a distinction to be able to define in the first case the abelian derived sequence. Let us start with a few definitions, similar to the ones given in Subsection 1.5.1.

Definition 3.3. Let u be a prefix of an abelian uniformly recurrent word $\mathbf{x}$. We say that a non-empty factor w of $\mathbf{x}$ is an *abelian return* to u, if there exists some $i \geq 0$ such that

- $\mathbf{x}[i, i+|w|-1] = w$,
- $\mathbf{x}[i, i+|u|-1] \sim_{ab} u \sim_{ab} \mathbf{x}[i+|w|, i+|w|+|u|-1]$,
- $\mathbf{x}[i+j, i+j+|u|-1] \not\sim_{ab} u$, for all $j \in \{1, \ldots, |w|-1\}$.

We denote by $\mathcal{APR}_{\mathbf{x},u}$ the set of abelian returns to the prefix u. Since $\mathbf{x}$ is abelian uniformly recurrent, then the set $\mathcal{APR}_{\mathbf{x},u}$ is finite. We define the set of abelian returns to prefixes as

$$\mathcal{APR}_{\mathbf{x}} := \bigcup_{u \in \mathrm{Pref}(\mathbf{x})} \mathcal{APR}_{\mathbf{x},u}.$$

Observe that if $\mathbf{x}$ is uniformly recurrent, then the length of the longest element in $\mathcal{APR}_{\mathbf{x},u}$ is bounded by the length of the longest element in $\mathcal{R}_{\mathbf{x},u}$.

We will also consider a more general situation where u is not restricted to be a prefix of $\mathbf{x}$. Puzynina and Zamboni [PZ13] called this notion a *semi-abelian return* to the abelian class of u and the number of abelian returns is the number of distinct abelian classes of semi-abelian returns.

Definition 3.4. If $\mathbf{x}$ is abelian recurrent and if u is a factor of $\mathbf{x}$, we can define as above the notion of abelian return to u. The corresponding set $\mathcal{AR}_{\mathbf{x},u}$ of abelian returns to u is well defined. We define the set of abelian returns as

$$\mathcal{AR}_{\mathbf{x}} := \bigcup_{u \in \mathrm{Fac}(\mathbf{x})} \mathcal{AR}_{\mathbf{x},u}.$$

Remark 3.5. Let $\mathbf{x}$ be an abelian recurrent word. The set $\mathcal{AR}_{\mathbf{x},u}$ is finite, for each factor u of $\mathbf{x}$, if and only if $\mathbf{x}$ is abelian uniformly recurrent.

3.2 Finiteness of the set of abelian returns

Puzynina and Zamboni [PZ13] provided a discussion between periodicity and the number of abelian returns and they give a sufficient condition for periodicity (see Lemma 3.2). Moreover they observe that a characterization of periodicity similar to Proposition 1.60 in terms of abelian returns does not exist. In the case of abelian returns, such a characterization does not hold in both direction. They provide the following counter-example.

Example 3.6 ([PZ13])**.** Consider an aperiodic word that is the concatenation of words of the form $\mathtt{110010}$ and $\mathtt{110100}$. Then, the factor $\mathtt{11}$ has exactly one abelian return which is: $\mathtt{110010} \sim_{ab} \mathtt{110100}$. So the existence of a factor having exactly one abelian return word does not guarantee the periodicity.

Now consider the periodic word $\mathbf{w} = (\mathtt{001101001011001100110011})^\omega$ of period 24. One can check that each factor of $\mathbf{w}$ has at least two abelian returns[1]. Hence, periodicity does not imply the existence of a factor having exactly one abelian return.

Here we take the finiteness of the set of abelian returns to characterize periodicity.

Theorem 3.7.

Let $\mathbf{x}$ be a recurrent word. The set $\mathcal{AR}_\mathbf{x}$ is finite if and only if $\mathbf{x}$ is periodic.

Proof. The “if” part is obvious. We prove the “only if” part.

Suppose that $\mathcal{AR}_\mathbf{x}$ is finite and that $\mathbf{x}$ is recurrent but not periodic. In this case, for each k, there exists a word u satisfying $|u| > k$ such that $au, bu \in \mathrm{Fac}(\mathbf{x})$ for some letters $a \neq b$. Hence there exist i, j such that $i < j$, $\mathbf{x}[i, i+|u|] = au$ and $\mathbf{x}[j, j+|u|] = bu$. Define $v = \mathbf{x}[i, j-1]$. Since $\mathbf{x}[i+d, j-1+d] \not\sim_{ab} v$ for all $d \in \{1, \ldots, |u|\}$, there is an abelian return to v in $\mathbf{x}$ of length at least k. As we can do the same for arbitrarily large k, the set $\mathcal{AR}_\mathbf{x}$ is infinite. ☐

Obviously, uniform recurrence implies abelian uniform recurrence, but the converse is not true.

Proposition 3.8.

There exists an abelian uniformly recurrent word which is not uniformly recurrent.

Proof. Let $\mathbf{t} = t_0 t_1 \cdots = \mathtt{01101001}\cdots$ be the Thue–Morse word and σ be the Thue–Morse morphism, $\sigma(\mathbf{t}) = \mathbf{t}$, introduced in Example 1.11. Define the set $I = \{i_0 < i_1 < \ldots\}$ of all positions where an isolated $\mathtt{1}$ occurs. That is, for all n, we have $t_{i_n} = \mathtt{1}$ and $t_{i_n - 1} = t_{i_n + 1} = \mathtt{0}$. Moreover we set $J = \{i_{2^k} \mid k > 0\}$.

[1] It suffices to check that each factor of length at most 24 has at least two abelian returns. For a factor of length longer than 24, its abelian returns coincide with abelian returns of a factor of length less than 24.

Let $\mathbf{y} = y_0 y_1 \cdots$ be the word defined by $y_j = \mathtt{0}$, if $j \in J$, and $y_j = t_j$ otherwise. Define $\mathbf{x} = \sigma(\mathbf{y})$:

i	i_0 i_1 i_2 i_3 i_4 i_5 i_6 i_7 i_8
$\mathbf{t}$	$\mathtt{0110100110010110100101100110100110010110011010010110100}\cdots$
$\mathbf{y}$	$\mathtt{0110100110000110000101100110000110010110011010010110000}\cdots$
$\mathbf{x}$	$\mathtt{0110100110010110100101100110100110010110011010010110100}\cdots$

The word $\mathbf{x}$ coincides with $\mathbf{t}$ almost everywhere, except for the positions from the set $2J \cup (2J+1)$. Hence, each factor of the Thue–Morse word occurs in $\mathbf{x}$ uniformly, i.e., with bounded gaps. At the same time, the factor $\sigma(\mathtt{000})$ occurs in $\mathbf{x}$ with strictly growing gaps. Hence $\mathbf{x}$ is not uniformly recurrent.

Let us now prove that $\mathbf{x}$ is abelian uniformly recurrent. First we point out a property of the Thue–Morse word: for all $d > 0$ and all $a \in \{\mathtt{0}, \mathtt{1}\}$, there exists k such that $t_k = a \neq t_{k+d}$. This property follows from the well-known fact that the Thue–Morse word does not contain any constant infinite arithmetical subsequence [MSS11].

As $\mathbf{x}$ is abelian periodic (of period 2), the weight (i.e., the sum of digits) of each factor u of $\mathbf{x}$ of odd length is either $\frac{|u|+1}{2}$ or $\frac{|u|-1}{2}$. Note that $y_i = \mathtt{0}$ implies $|\mathbf{x}[2i+1, 2i+|u|]|_{\mathtt{1}} = \frac{|u|+1}{2}$ and $y_i = \mathtt{1}$ implies $|\mathbf{x}[2i+1, 2i+|u|]|_{\mathtt{1}} = \frac{|u|-1}{2}$. Since $\mathtt{0}$ (resp. $\mathtt{1}$) occurs with bounded gaps in $\mathbf{y}$, gaps between abelian occurrences in $\mathbf{x}$ of a factor of odd length are bounded.

The weight of a factor u of even length of $\mathbf{x}$ can take values $\frac{|u|}{2}$, $\frac{|u|}{2}+1$ and $\frac{|u|}{2}-1$. The first case takes place when u occurs at an even position in $\mathbf{x}$, meaning that the gaps between abelian occurrences of u of weight $\frac{|u|}{2}$ in $\mathbf{x}$ are bounded. The last two cases take place if u occurs in $\mathbf{x}$ at an odd position i and if $y_{\frac{i-1}{2}} = \mathtt{1}$ and $y_{\frac{i-1+|u|}{2}} = \mathtt{0}$ or, $y_{\frac{i-1}{2}} = \mathtt{0}$ and $y_{\frac{i-1+|u|}{2}} = \mathtt{1}$. Due to the mentioned property of the Thue–Morse word, there exists k such that $t_k = \mathtt{1} \neq t_{k+\frac{|u|}{2}}$ (resp. $t_k = \mathtt{0} \neq t_{k+\frac{|u|}{2}}$) and since $\mathbf{t}$ is uniformly recurrent, the factor $\mathbf{t}[k, k+\frac{|u|}{2}]$ occurs infinitely often with bounded gaps in $\mathbf{t}$. Hence abelian occurrences of u in $\mathbf{x}$ appear infinitely often with bounded gaps.

$\square$

3.3 Finiteness of the set of abelian returns to prefixes

Contrary to the finiteness of $\mathcal{AR}_{\mathbf{x}}$, the finiteness of $\mathcal{APR}_{\mathbf{x}}$ does not imply periodicity nor abelian periodicity of $\mathbf{x}$. Moreover, if $\mathbf{x}$ is uniformly recurrent, it is well-known that

$$\min_{v \in \mathcal{R}_{\mathbf{x},u}} |v| \to \infty, \text{ if } |u| \to \infty,$$

meaning that taking longer prefixes eventually leads to longer return words. Here we show that such a result does not hold for abelian returns to prefixes. Indeed, for the Thue–Morse word the corresponding set $\mathcal{APR}_{\mathbf{t}}$ is finite and can be described precisely. Such a result also holds for the Fibonacci word. In particular, amongst the set of Sturmian words, the finiteness of $\mathcal{APR}_{\mathbf{x}}$ characterizes Sturmian words with non-zero intercept.

Lemma 3.9.

If $\mathbf{x}$ is a uniformly recurrent and abelian periodic word, then the set $\mathcal{APR}_{\mathbf{x}}$ is finite.

Proof. Let m be the (minimal) abelian period of $\mathbf{x}$. Let us find an upper bound on the length of an abelian return u to a prefix p of $\mathbf{x}$.

Suppose first that $|p| = mk$. In this case, due to abelian periodicity, for all i, we have $\mathbf{x}[mi, m(i+k)-1] \sim_{ab} p$. Hence we get $|u| \leqslant m$.

Suppose now that $|p| = mk+\ell$, with $0 < \ell < m$. Let us denote the word $\mathbf{x}[mk, m(k+1)-1]$ by s. If there exists i such that the equality $\mathbf{x}[mi, m(i+1)-1] = s$ holds, then we have $\mathbf{x}[m(i-k), mi+\ell-1] \sim_{ab} p$ as the word $\mathbf{x}$ is abelian periodic. Hence, it is sufficient to prove that the set

$$\{i \geq 0 \mid \mathbf{x}[mi, m(i+1)-1] = s\}$$

has bounded gaps.

Let us consider the word $\mathbf{x}'$ over the alphabet of factors of $\mathbf{x}$ of length m, such that $\mathbf{x}'_i = \mathbf{x}[mi, m(i+1)-1]$. It is well-known that the uniform recurrence of $\mathbf{x}$ implies uniform recurrence of $\mathbf{x}'$ (see for instance [Sal10]). Hence, for each letter of $\mathbf{x}'$ there is an upper bound on the gap between two consecutive occurrences of it in $\mathbf{x}'$. Denoting the maximum of such constants by D, we get $|u| \leqslant mD$. $\square$

Remark 3.10. In Lemma 3.9, the condition on a word $\mathbf{x}$ to be uniformly recurrent is essential: there exists an abelian periodic word $\mathbf{x}$ which is not uniformly recurrent and such that $\mathcal{APR}_{\mathbf{x},u}$ is infinite for some prefix u of $\mathbf{x}$. Consider the abelian periodic word of period 4 given by $\mathbf{x} = \phi\varphi^{\omega}(\mathtt{0})$ where $\varphi : \mathtt{0} \mapsto \mathtt{010}, \mathtt{1} \mapsto \mathtt{111}$ and $\phi : \mathtt{0} \mapsto \mathtt{01230123}, \mathtt{1} \mapsto \mathtt{0213}$:

$$\mathbf{x} = \mathtt{01230123\ \ 0213\ \ 01230123\ \ 0213\ \ 0213\ \ 0213} \cdots$$

In $\mathbf{x}$ there are unbounded gaps between consecutive abelian occurrences of its prefix 012301 that correspond to the occurrences of $\phi(\mathtt{1}^m)$.

Remark 3.11. In Lemma 3.9, the condition on abelian periodicity of $\mathbf{x}$ is not necessary to get finiteness of $\mathcal{APR}_{\mathbf{x}}$. We shall give an example below when discussing the case of Sturmian words. Indeed, Sturmian words are not abelian periodic (see Lemma 3.19) but for instance, the Fibonacci word $\mathbf{f}$ is uniformly recurrent and the corresponding set $\mathcal{APR}_{\mathbf{f}}$ is finite.

Proposition 3.12.

A word $\mathbf{x}$ is periodic if and only if there exists some prefix u such that infinitely many factors of $\mathbf{x}$ are abelian equivalent to u and all the abelian returns in $\mathcal{APR}_{\mathbf{x},u}$ have length 1.

Proof. If $\mathbf{x} = u^{\omega}$, then $\mathbf{x}[i, i+|u|-1] \sim_{ab} u$ for all $i \geq 0$. Conversely, if all the abelian returns to some prefix u in $\mathcal{APR}_{\mathbf{x},u}$ have length 1, then $\mathbf{x}[i, i+|u|-1] \sim_{ab} u \sim_{ab} \mathbf{x}[i+1, i+|u|]$ for all $i \geq 0$. There is an abelian return a of length 1 at position i in $\mathbf{x}$ and it also occurs in position $i+|u|$. It follows that $|u|$ is a period of $\mathbf{x}$. $\square$

3.3.1 The set $\mathcal{APR}_{\mathbf{t}}$ for the Thue–Morse word $\mathbf{t}$ is finite

We already know from Lemma 3.9 that the Thue–Morse word has a finite set of abelian returns to all its prefixes. Here we describe precisely this set.

Lemma 3.13.

Let $\mathbf{x}$ be a uniformly recurrent word. Let $n \geq 1$ and i, j be such that $i < j$. Assume that $\mathbf{x}[i, i+n-1] \sim_{ab} \mathbf{x}[j, j+n-1]$ and there exists a prefix u of length $j-i$ of $\mathbf{x}$ such that $u \sim_{ab} \mathbf{x}[i, j-1]$. The word $\mathbf{x}[i, i+n-1]$ is an occurrence of an abelian return to the prefix u if and only if, for all $\ell \in \{0, \ldots, n-2\}$, $\mathbf{x}[i, i+\ell] \not\sim_{ab} \mathbf{x}[j, j+\ell]$.

Proof. Since $|u| = j - i$, by assumption we have $\mathbf{x}[i, i+|u|-1] \sim_{ab} u$. Observe first that there exists $\ell \in \{0, \ldots, n-1\}$ such that $\mathbf{x}[i, i+\ell] \sim_{ab} \mathbf{x}[j, j+\ell]$ if and only if $\mathbf{x}[i+\ell+1, i+\ell+|u|] \sim_{ab} u$. In particular, since $\mathbf{x}[i, i+n-1] \sim_{ab} \mathbf{x}[j, j+n-1]$, we get $\mathbf{x}[i+n, i+n+|u|-1] = \mathbf{x}[i+n, j-1+n] \sim_{ab} u$. Moreover, $\ell \in \{0, \ldots, n-2\}$ is such that $\mathbf{x}[i+\ell+1, i+\ell+|u|] \not\sim_{ab} u$ if and only if $\mathbf{x}[i, i+\ell] \not\sim_{ab} \mathbf{x}[j, j+\ell]$. □

Remark 3.14. From this lemma, we can derive a necessary condition for a word to be an abelian return to a prefix. If a word $w = w_1 \cdots w_n$ of length n is an abelian return to a prefix, then there exists some factor $y = y_1 \cdots y_n$ of $\mathbf{x}$ such that

$$w \sim_{ab} y \text{ and, for all } \ell \in \{1, \ldots, n-1\}, w_1 \cdots w_\ell \not\sim_{ab} y_1 \cdots y_\ell. \tag{3.1}$$

This condition is not sufficient. For instance, $w = \texttt{001011}$ and $y = \texttt{110010}$ are two factors of length 6 satisfying(3.1) and occurring in the Thue–Morse word $\mathbf{t}$. But, as shown in the following proposition, w is not an abelian return to any prefix.

Theorem 3.15.

The set $\mathcal{APR}_{\mathbf{t}}$ of abelian returns to prefixes for the Thue–Morse word $\mathbf{t}$ is

$$\{\texttt{0, 1, 01, 10, 001, 011, 100, 110, 0011, 0101,}$$
$$\texttt{1010, 1100, 00101, 01011, 10100, 11010}\}.$$

Proof. One can check with some computer experiments that the factors given above appear as abelian returns to some prefix of $\mathbf{t}$. Moreover, one can also check that these are the only factors of length $2, \ldots, 5$ in $\mathbf{t}$ satisfying condition (3.1).

Assume that there exists some abelian return $w = w_1 \cdots w_n = \mathbf{t}[i, i+n-1]$ of length $n \geq 2$ to a prefix of $\mathbf{t}$ occurring at position i. In particular, we may assume that w is an abelian return to the prefix u of length $j - i > 0$ and $y = y_1 \cdots y_n = \mathbf{t}[j, j+n-1]$ satisfies (3.1). We will show that the length of w is at most 5. Recall that $\mathbf{t}[2k, 2k+1] \in \{\texttt{01}, \texttt{10}\}$ for all $k \geq 0$.

Assume first that i, j are even. Since $\mathbf{t}[i, i+1]$ and $\mathbf{t}[j, j+1]$ belong to $\{\texttt{01}, \texttt{10}\}$, we conclude that $\mathbf{t}[i, i+1] \sim_{ab} \mathbf{t}[j, j+1]$ and, in that situation, we can only have an abelian return of length at most 2.

Assume now that i is odd and j is even and that $\mathbf{t}_i = \texttt{0}$ (symmetric cases can be treated in the same way). Our aim is to build the longest possible abelian return. Since $\mathbf{t}_i = \texttt{0}$ and j is even, we consider $\mathbf{t}[j, j+1] = \texttt{10}$ because otherwise, $\mathbf{t}[j, j+1] = \texttt{01}$ and $w_1 = y_1$ (i.e., $\mathbf{t}[i+1, j+1] \sim_{ab} \mathbf{t}[i, j] \sim_{ab} u$ and we get directly an abelian return of length $n = 1$). Now $\mathbf{t}[i, i+2] = \texttt{001}$ because otherwise, $\mathbf{t}[i, i+2] = \texttt{010}$ and $w_1 w_2 \sim_{ab} y_1 y_2$. Continuing this way, we have $\mathbf{t}[j, j+3] = \mathtt{1010}$ and $\mathbf{t}[i, i+4] = \mathtt{00101}$. Since $(\mathtt{10})^3$ is not a factor of $\mathbf{t}$, we have $\mathbf{t}[j, j+5] = \mathtt{101001}$ and $\mathbf{t}[i, i+4] \sim_{ab} \mathbf{t}[j, j+4]$. In that situation, we can only have an abelian return of length at most 5.

The last case is when i and j are odd. Assume $\mathbf{t}_i = \mathtt{0}$ and $\mathbf{t}_j = \mathtt{1}$. We have $\mathbf{t}_i = \mathtt{0}$ and $\mathbf{t}_{j-1} = \mathtt{0}$ because $\mathbf{t}[j-1, j] = \mathtt{01}$. Moreover, $z = \mathbf{t}[i+1, j-2] \in \{\mathtt{01}, \mathtt{10}\}^*$ and thus $v = \mathbf{t}[i, j-1] = \mathtt{0}z\mathtt{0}$ is a word of even length such that $|v|_\mathtt{0} = 2 + |v|_\mathtt{1}$. Therefore v cannot be abelian equivalent to a prefix u of $\mathbf{t}$. So in such a situation, we cannot have an abelian return to some prefix of $\mathbf{t}$. □

Proposition 3.16.

If a factor of length $n \geq 6$ of the Thue–Morse word satisfies (3.1), then n is even.

Proof. Let $w = \mathbf{t}[i, i+n-1]$ and $y = \mathbf{t}[j, j+n-1]$ be factors of $\mathbf{t}$ of length $n \geq 6$ satisfying (3.1). As $n \geq 6$, i and j are odd. Hence, to satisfy the condition (3.1), we must have

$$\begin{pmatrix} \mathbf{t}[i, i+n-1] \\ \mathbf{t}[j, j+n-1] \end{pmatrix} \in \begin{pmatrix} \mathtt{0} \\ \mathtt{1} \end{pmatrix} \left\{ \begin{pmatrix} \mathtt{01} \\ \mathtt{01} \end{pmatrix}, \begin{pmatrix} \mathtt{10} \\ \mathtt{10} \end{pmatrix}, \begin{pmatrix} \mathtt{01} \\ \mathtt{10} \end{pmatrix} \right\}^* \begin{pmatrix} \mathtt{1} \\ \mathtt{0} \end{pmatrix}$$
$$\bigcup \begin{pmatrix} \mathtt{1} \\ \mathtt{0} \end{pmatrix} \left\{ \begin{pmatrix} \mathtt{01} \\ \mathtt{01} \end{pmatrix}, \begin{pmatrix} \mathtt{10} \\ \mathtt{10} \end{pmatrix}, \begin{pmatrix} \mathtt{10} \\ \mathtt{01} \end{pmatrix} \right\}^* \begin{pmatrix} \mathtt{0} \\ \mathtt{1} \end{pmatrix}.$$

So n must be even. □

For $n = 6, 8, 10, \ldots, 104$, with a computer search, we get the following number of factors of length n satisfying (3.1): 6, 4, 8, 12, 12, 4, 8, 8, 4, 0, 0, 8, 0, 0, 4, 8, 4, 0, 0, 0, 0, 0, 4, 0, 4, 0, 0, 0, 0, 0, 4, 8, 4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 8, 0, 0.

3.3.2 The finiteness of $\mathcal{APR}_\mathbf{x}$ for a Sturmian word x depends on its intercept

Let $\mathbf{x} = St(\alpha, \rho)$ be a Sturmian word. For a binary word $v = v_0 v_1 \cdots v_m$, we define a half-interval I_v of C as

$$I_v := I_{v_0} \cap R_\alpha^{-1}(I_{v_1}) \cap \cdots \cap R_\alpha^{-m}(I_{v_m}). \tag{3.2}$$

Hence $\mathbf{x}[i, i+m] = v$ if and only if $R_\alpha^i(\rho) \in I_v$. See [Lot02, Section 2.1.2].

Definition 3.17. Let $\mathbf{x} = St(\alpha, \rho)$ be a Sturmian word. For each k the number of $\mathtt{1}$'s in a factor of length k in $\mathbf{x}$ can only take the values $\lceil k\alpha \rceil$ or $\lceil k\alpha \rceil - 1$. The corresponding factors will be called respectively *heavy* and *light*. If $\mathbf{x}$ is understood from the context, $H(k)$ (resp. $L(k)$) will denote the set of heavy (resp. light) factors of length k in $\mathbf{x}$. Define

$$I_H(k) := \bigcup_{v \in H(k)} I_v \text{ and } I_L(k) := \bigcup_{v \in L(k)} I_v.$$

So, the word $\mathbf{x}[i, i+k-1]$ is heavy if and only if $R_\alpha^i(\rho) \in I_H(k)$.

Theorem 3.18.

Let $\mathbf{x}$ be a Sturmian word. The set $\mathcal{APR}_{\mathbf{x}}$ is finite if and only if $\mathbf{x}$ does not have a null intercept.

Proof. For the sake of convenience, let $\mathbf{x}$ be defined as $St(\alpha,\rho)$ for half-intervals $I_0 = [0, 1-\alpha)$ and $I_1 = [1-\alpha, 1)$. Let us prove by induction on $k \geq 1$ that

$$I_H(k) = [1 - \{k\alpha\}, 1) \text{ and } I_L(k) = [0, 1 - \{k\alpha\}). \tag{3.3}$$

It holds true for $k = 1$. Suppose now that the statement holds true for some $k \geq 1$. We consider two cases.

- Assume that $0 \notin R_\alpha^{-k}(I_1)$. Therefore we get $R_\alpha^{-k}(I_1) = [1 - \{(k+1)\alpha\}, 1 - \{k\alpha\})$ with $1 - \{(k+1)\alpha\} < 1 - \{k\alpha\}$. By the inductive hypothesis, we have $I_H(k) = [1 - \{k\alpha\}, 1)$ and consequently,
$$R_\alpha^{-k}(I_1) \cap I_H(k) = \emptyset.$$
This means that all the heavy factors of length k of $\mathbf{x}$ can only be extended with `0` to factors of length $k+1$ of $\mathbf{x}$. In particular, the weights of heavy factors of length k and $k+1$ are the same. At the same time, we have $R_\alpha^{-k}(I_1) \cap I_L(k) = R_\alpha^{-k}(I_1)$, which means that the factors corresponding to elements belonging to this latter set are the light factors of length k that are extended with `1` to heavy factors of length $k+1$. We conclude that
$$I_H(k+1) = I_H(k) \cup R_\alpha^{-k}(I_1) = [1 - \{(k+1)\alpha\}, 1)$$
and $I_L(k+1) = I_L(k) \backslash R_\alpha^{-k}(I_1) = [0, 1 - \{(k+1)\alpha\})$.
- Assume now that $0 \in R_\alpha^{-k}(I_1)$, i.e., $1 - \{(k+1)\alpha\} > 1 - \{k\alpha\}$. In this case, using again the inductive hypothesis, $R_\alpha^{-k}(I_1) \cap I_H(k) = [1 - \{(k+1)\alpha\}, 1)$ is non-empty. This interval corresponds to the heavy factors of length k having an extension with `1` making them the only heavy factors of length $k+1$ in $\mathbf{x}$.

Now we are ready to prove the main part of the statement. First of all, let us prove that, if $\mathbf{x}$ has a null intercept, then $\mathcal{APR}_{\mathbf{x}}$ is infinite. Let $k \geq 1$ and p be the prefix of length k of $\mathbf{x}$. As $\rho = 0$, we have $0 \in I_p$. Since the interval I_p corresponds exactly to one word p which is either light or heavy, we have $I_p \subseteq I_L(k)$ or $I_p \subseteq I_H(k)$. As $0 \in I_p$, we conclude that $I_p \subseteq I_L(k)$ using (3.3). In other words, we have just shown that each prefix of $\mathbf{x}$ is a light factor.

Now we show that, for all n, there exists a length ℓ such that gaps between consecutive occurrences of light factors of length ℓ in $\mathbf{x}$ can be larger than n. Let $n \geq 1$. Define the set of points
$$S_n := \{R_\alpha^i(0) \mid 0 \leqslant i \leqslant n\}$$
and denote by d the minimal length of intervals having endpoints in S_n. Due to Kronecker's theorem, we can find some ℓ such that $|I_L(\ell)| < d$ and it follows that $I_L(\ell) \cap S_n = \{0\}$. With our definitions, it means that the light prefix of $\mathbf{x}$ of length ℓ is followed by at least n heavy consecutive factors of length ℓ. Since this can be done for any n, the set $\mathcal{APR}_{\mathbf{x}}$ is infinite.

Let us prove that, if $\mathbf{x}$ does not have a null intercept, then $\mathcal{APR}_{\mathbf{x}}$ is finite. The main difference with the previous situation is that a prefix can now be a heavy or a light factor depending on its length: $\rho \geq 1 - \{k\alpha\}$ if and only the prefix of length k is heavy. We will show that there exists a constant c such that, for all prefixes p of $\mathbf{x}$, the gap between consecutive occurrences of factors abelian equivalent to p is bounded by c.

Let $n \geq 1$. Consider as before the set S_n and order its elements $0 = s_0 < s_1 < \cdots < s_n$. Denote by $D(n)$ the maximal length of the intervals $[s_0, s_1)$, $\ldots$, $[s_{n-1}, s_n)$, $[s_n, s_0)$ whose endpoints are consecutive points in S_n. Due to Kronecker's theorem, there exists some c such that $2D(c) < \min\{\rho, 1-\rho\}$.

Suppose that the prefix of length k of $\mathbf{x}$ is a light word. Then we have $\rho \in I_L(k)$ and, consequently, $|I_L(k)| > \rho$. Assume that there is a light factor of length k occurring at position i in $\mathbf{x}$, i.e., $R^i_\alpha(\rho) \in I_L(k)$. We consider two cases. If $R^i_\alpha(\rho) \geq D(c)$, there exists $j \in \{1, \ldots, c\}$ such that $R^j_\alpha(0) = s_c$ and $\theta = 1 - R^j_\alpha(0) \in (0, D(c)]$. Hence the point $R^j_\alpha(R^i_\alpha(\rho)) = R^{i+j}_\alpha(\rho) = R^i_\alpha(\rho) - \theta$ belongs to $I_L(k)$, i.e., the factor of length k at position $i + j$ in $\mathbf{x}$ is light again. If $R^i_\alpha(\rho) < D(c)$, there exists $j \in \{1, \ldots, c\}$ such that $R^j_\alpha(0) = s_1 \leq D(c) < \rho/2$. Hence the point $R^j_\alpha(R^i_\alpha(\rho)) = R^{i+j}_\alpha(\rho) = R^i_\alpha(\rho) + s_1$ is less than ρ and belongs to $I_L(k)$.

A similar proof can be done for the case of a heavy prefix of length k. Assume that $\rho \in I_H(k)$ and that, for some $i \geq 0$, $R^i_\alpha(\rho) \in I_H(k)$. If $R^i_\alpha(\rho) < 1-D(c)$, then $R^i_\alpha(\rho)+s_1 < 1$. If $R^i_\alpha(\rho) \geq 1-D(c)$, then $R^i_\alpha(\rho)-(1-s_c) \geq 1-2D(c) > \rho$. We can derive the same conclusion as above.

Hence the number c is an upper bound on the length of abelian returns to any prefix and therefore $\mathcal{APR}_{\mathbf{x}}$ is finite. $\square$

Lemma 3.19.

No Sturmian word is abelian periodic.

Proof. Proceed by contradiction and assume that $\mathbf{x} = St(\alpha, \rho)$ is abelian periodic of period m with α irrational. Then all factors of the kind $\mathbf{x}[tm, (t+1)m-1]$, $t \in \mathbb{N}$, are abelian equivalent, i.e., have the same weight. Assume that, for all t, $R^{tm}_\alpha(\rho) = R^t_{m\alpha}(\rho) \in I_L(m)$. But since α is irrational, $m\alpha$ is also irrational and thanks to Kronecker's theorem, $\{R^t_{m\alpha}(\rho) \mid t \geq 0\}$ is dense in C contradicting the fact that $\{R^t_{m\alpha}(\rho) \mid t \geq 0\} \cap I_H(m)$ should be empty. $\square$

3.3.3 The set $\mathcal{APR}_{\mathbf{f}}$ for the Fibonacci word f is finite

From Theorem 3.18, since the Fibonacci word $\mathbf{f}$ is given by $St(1/\phi^2, 1/\phi^2)$, we already know that $\mathcal{APR}_{\mathbf{f}}$ is finite. Here we exhibit exactly the elements of this set in Theorem 3.21. As a first attempt, (3.1) gives a necessary condition that allows one to exclude some words as abelian returns. This condition will not be used in the proof of Theorem 3.21 but, interestingly, our developments can be related to the lengths of the palindromic prefixes of $\mathbf{f}$, [dL97, Fis06].

Proposition 3.20.

For the Fibonacci word, there exist exactly two factors of length n satisfying (3.1) if n is a Fibonacci number. Otherwise, no factor of length n satisfies such a condition.

Proof. Consider two factors x, y of length n satisfying (3.1) and occurring in the Fibonacci word $\mathbf{f}$. Assume that x starts with $\mathtt{0}$. Then to fulfill (3.1), y starts with $\mathtt{1}$. Since $\mathbf{f}$ is Sturmian, for any two words of the same length x' and y' which are prefixes of x and y respectively, we have $||x'|_\mathtt{1} - |y'|_\mathtt{1}| \leq 1$. Therefore, we deduce that x and y must be of the form $x = \mathtt{0}u\mathtt{1}$ and $y = \mathtt{1}u\mathtt{0}$ for some $u \in \{\mathtt{0}, \mathtt{1}\}^*$. This means that u is a bispecial factor of the Fibonacci word.

Recall that the left special factors in $\mathbf{f}$ are its prefixes and its right special factors are the mirror images of its prefixes [CN10, Prop. 4.10.3]. So bispecial factors of $\mathbf{f}$ are its palindromic prefixes. If $(\ell_i)_{i\geq 1}$ denotes the increasing sequence of all lengths of palindromic prefixes in $\mathbf{f}$, it is well-known that $(\ell_i)_{i\geq 1} = (0, 1, 3, 6, 11, \ldots)$ is given by $\ell_i = F_{i+1} - 2$ where F_i is the ith Fibonacci number. See [dL97, Theorem 5] and [Fis06]. Hence n must be a Fibonacci number.

Conversely, for any bispecial factor u of $\mathbf{f}$, it is easy to show that either $\mathtt{0}u\mathtt{0}$ or $\mathtt{1}u\mathtt{1}$ is not a factor occurring in $\mathbf{f}$ (see for instance [Lot02, p. 47]). Therefore, amongst the four words $\mathtt{0}u\mathtt{0}$, $\mathtt{1}u\mathtt{1}$, $\mathtt{0}u\mathtt{1}$ and $\mathtt{1}u\mathtt{0}$, the last two must occur in $\mathbf{f}$ and we get exactly two factors of length $|u| + 2$ satisfying (3.1). Indeed, assume that $\mathtt{0}u\mathtt{0}$ does not occur in $\mathbf{f}$. Then for u to be left (resp. right) special, $\mathtt{0}u\mathtt{1}$ (resp. $\mathtt{1}u\mathtt{0}$) must occur in $\mathbf{f}$. □

The reader may notice that the computations carried out in the proofs of the next two results could also be adapted to other Sturmian words.

Theorem 3.21.

The set $\mathcal{APR}_\mathbf{f}$ of abelian returns to prefixes for the Fibonacci word $\mathbf{f}$ contains exactly the words `0`, `1`, `01`, `10`, `001`.

Proof. Using the same notation as in Theorem 3.18, for $c = 7$, we have $D(7) \approx 0.145898$ which is such that $2D(7) < \min\{1/\phi^2, 1 - 1/\phi^2\} \approx 0.381$. Hence, all abelian returns to prefixes of the Fibonacci word have length at most 7. Actually, this value can be reduced.

Lemma 3.22.

There is no abelian return of length greater than 3 to prefixes in the Fibonacci word.

Proof. With the notation of the proof of Theorem 3.18, we set $\rho = \alpha = 1/\phi^2 \approx 0.381$. Let i be a natural number. Define the four points $\rho_{i,t} = R_\alpha^{i+t}(\rho)$ for $t = 0, 1, 2, 3$. Recall that, for all $k \geq 1$, the unit circle $[0, 1)$ is split into two half-intervals $I_H(k) = [1 - \{k\alpha\}, 1)$ and $I_L(k) = [0, 1 - \{k\alpha\})$ such that two factors $\mathbf{f}[i, i + k - 1]$ and $\mathbf{f}[j, j + k - 1]$ are abelian equivalent if and only if the points $R^i_\alpha(\rho)$ and $R^j_\alpha(\rho)$ belong to the same interval $I_H(k)$ or $I_L(k)$.

Let I be any of the two intervals $I_H(k)$ or $I_L(k)$. What we are going to prove is that if ρ and $\rho_{i,0}$ belong to I, then either $\rho_{i,2}$ or $\rho_{i,3}$ belongs to I. In other words, if we have $\mathbf{f}[i, i+k-1] \sim_{ab} \mathbf{f}[0, k-1]$, then either $\mathbf{f}[i+2, i+k+1]$ or $\mathbf{f}[i+3, i+k+2]$ is abelian equivalent to $\mathbf{f}[i, i+k-1]$ which gives the upper bound on the length of any abelian return to a prefix of $\mathbf{f}$.

Note that $\rho_{i,0} = R_\delta(\rho_{i,2})$, $\rho_{i,2} = R_{-\delta}(\rho_{i,0})$ and $\rho_{i,3} = R_{\alpha-\delta}(\rho_{i,0})$, where δ is equal to $1-2\alpha \approx 0.2361$. Assume that the factor of length k starting in position i is abelian equivalent to the prefix of $\mathbf{f}$ of length k, i.e., ρ and $\rho_{i,0}$ are both light or heavy words. We consider two cases. Suppose first that $\rho, \rho_{i,0} \in I_L(k)$. In this case, we have $[0, \rho] \subseteq I_L(k)$.

- If $\rho \leq \rho_{i,0} < 1$, then $[0, \rho_{i,0}] \subseteq I_L(k)$ and $0 < \rho_{i,2} = R_{-\delta}(\rho_{i,0}) < \rho_{i,0}$. Thus $\rho_{i,2}$ belongs also to $I_L(k)$.
- If $\rho > \rho_{i,0} > 0$, either $\rho_{i,0} \geq \delta$ and then $\rho_{i,2} = R_{-\delta}(\rho_{i,0}) \in [0, \rho)$ meaning that $\rho_{i,2}$ is in $I_L(k)$ or, $0 < \rho_{i,0} < \delta$, i.e. $-\delta < \rho_{i,2} - 1 < 0$ and then $0 < \alpha - \delta < \rho_{i,3} = R_\alpha(\rho_{i,2}) < \rho$ meaning that $\rho_{i,3} \in I_L(k)$.

Suppose now that $\rho, \rho_{i,0} \in I_H(k)$. In this case, as $\rho \in I_H(k)$, we have $[\rho, 1) \subseteq I_H(k)$.

- If $\rho > \rho_{i,0} > 0$, then $[\rho_{i,0}, 1) \subseteq I_H(k)$ and $\rho_{i,3} = R_{\alpha-\delta}(\rho_{i,0})$ belongs to $I_H(k)$.
- If $\rho \leq \rho_{i,0} < 1$, either $\rho_{i,0} \geq \rho+\delta$ and then $\rho_{i,2} = R_{-\delta}(\rho_{i,0}) \in I_H(k)$ or, $\rho \leq \rho_{i,0} < \rho+\delta$, i.e., $\rho - \delta \leq \rho_{i,2} < \rho$ and then $\rho < \rho - \delta + \alpha \leq \rho_{i,3} = R_\alpha(\rho_{i,2}) < \rho + \alpha < 1$ meaning that $\rho_{i,3} \in I_H(k)$.

□

The factors of length at most 3 occurring in $\mathbf{f}$ are ε, `0`, `1`, `00`, `01`, `10`, `001`, `010`, `100` and `101`. Clearly, `00`, `010` and `101` do not satisfy (3.1) and cannot be abelian returns. To conclude the proof, we just have to show that `100` is also forbidden.

Lemma 3.23.

The set $\mathcal{APR}_\mathbf{f}$ of abelian returns to prefixes for the Fibonacci word $\mathbf{f}$ does not contain `100`.

Proof. We continue with notation of Lemma 3.22. Suppose that $\mathtt{100} \in \mathcal{APR}_\mathbf{f}$. There exists a prefix p of $\mathbf{f}$ of length k and a position $i \geq 0$ such that

1. $\mathbf{f}[i, i+2] = \mathtt{100}$,
2. $\mathbf{f}[i, i+k-1] \sim_{ab} p$, i.e., ρ and $\rho_{i,0}$ belong to the same interval $I \in \{I_L(k), I_H(k)\}$,
3. for $t = 1, 2$, $\mathbf{f}[i+t, i+t+k-1] \not\sim_{ab} p$, i.e., $\rho_{i,1}$ and $\rho_{i,2}$ do not belong to I,
4. $\mathbf{f}[i+3, i+2+k] \sim_{ab} p$.

To get a contradiction, let us prove that either $\rho_{i,1}$ or $\rho_{i,2}$ belongs to I. Since $\mathbf{f}_i = \mathtt{1}$, $\rho_{i,0}$ belongs to $I_1 = [1-\alpha, 1)$. If $I = I_L(k)$, then we have $\rho_{i,1} \in [0, \rho) \subseteq I_L(k)$. If $I = I_H(k)$, then we have $\rho_{i,2} \in [\rho, \rho+\alpha) \subseteq I_H(k)$. □

That concludes the proof of Theorem 3.21. □

3.3.4 The finiteness of $\mathcal{APR}_{\mathbf{x}}$ implies a bounded abelian complexity of $\mathbf{x}$

We show that the finiteness of $\mathcal{APR}_{\mathbf{x}}$ implies a bounded abelian complexity of $\mathbf{x}$, but the converse does not hold in general. Indeed, the abelian complexity of any Sturmian word $\mathbf{x}$ satisfies $\mathcal{P}^{(1)}_{\mathbf{x}}(n) = 2$ for all $n \geq 1$ (Equation (1.1)): there are exactly two kinds of factors of length n, the light ones and the heavy ones. But thanks to Theorem 3.18, if $\mathbf{x}$ is a Sturmian word with null intercept, then $\mathcal{APR}_{\mathbf{x}}$ is infinite. In other words, bounded abelian complexity does not imply the finiteness of $\mathcal{APR}_{\mathbf{x}}$.

Proposition 3.24.

If $\mathbf{x}$ is an abelian recurrent word such that $\mathcal{APR}_{\mathbf{x}}$ is finite, then $\mathbf{x}$ has bounded abelian complexity.

Proof. Suppose $\mathbf{x}$ satisfies the assumptions of the proposition but that $\mathbf{x}$ has unbounded abelian complexity. From Lemma 1.40, we deduce that there exists a symbol a such that the maximum of differences $|u|_a - |v|_a$ for factors u, v in $\mathbf{x}$ having equal length can be arbitrarily large.

Let $\delta > 0$. There exist $u, v \in \mathrm{Fac}(\mathbf{x})$ of equal length n such that $|u|_a - |v|_a \geq \delta$. Let $p = x_0 x_1 \ldots x_{n-1}$ be the prefix of length n of $\mathbf{x}$. Without loss of generality, we may assume that

$$||u|_a - |p|_a| \geqslant \frac{\delta}{2}.$$

Indeed, if $||u|_a - |p|_a| < \delta/2$ and $||v|_a - |p|_a| < \delta/2$, then one would deduce that $||u|_a - |v|_a| < \delta$.

As $\mathbf{x}$ is abelian recurrent, factors abelian equivalent to p (resp. to u) occur infinitely often in $\mathbf{x}$. Therefore there exist $i < j < k$ such that

1. $\mathbf{x}[i, i+n-1] \sim_{ab} p$, $\mathbf{x}[k, k+n-1] \sim_{ab} p$,
2. for all t such that $i < t < k$, we have $\mathbf{x}[t, t+n-1] \not\sim_{ab} p$,
3. $\mathbf{x}[j, j+n-1] \sim_{ab} u$.

This just means that we can consider two consecutive factors abelian equivalent to p separated by a factor abelian equivalent to u. Note that, for all t,

$$||x[t+c, t+n-1+c]|_a - |x[t, t+n-1]|_a| \leqslant c, \quad \forall c \leq n.$$

Hence, $j - i \geq \delta/2$ and $k - j \geqslant \delta/2$. Therefore we get $k - i \geq \delta$ which means that the abelian return $\mathbf{x}[i, k-1]$ to the prefix p has length at least δ. As δ can be chosen arbitrarily large, the set $\mathcal{APR}_{\mathbf{x}}$ is infinite and that is a contradiction. □

3.4 Abelian derived sequences

We refer the reader to definitions and notation introduced in Section 3.1. As was studied by Durand [Dur98] for classical return words, we introduce the notion of abelian derived sequence which is the factorization of an infinite word with respect to its abelian returns to prefixes in their order of occurrence. The next result allows us to define such a sequence.

Lemma 3.25.

Let u be a prefix of a uniformly recurrent word $\mathbf{x}$. The word $\mathbf{x}$ has a factorization as a sequence $m_0m_1m_2\cdots$ of elements in $\mathcal{APR}_{\mathbf{x},u}$ computed as follows. Consider the sequence of indices $(i_n)_{n\geq 0}$ such that, for all $j\geq 0$, $\mathbf{x}[i_j, i_j+|u|-1] \sim_{ab} u$ and, for all $i \notin \{i_n \mid n\geq 0\}$, we have $\mathbf{x}[i, i+|u|-1] \not\sim_{ab} u$. Set $m_n := \mathbf{x}[i_n, i_{n+1}-1]$.

As shown in Example 3.27, the factorization of $\mathbf{x}$ with elements in $\mathcal{APR}_{\mathbf{x},u}$ is not necessarily unique.

Definition 3.26. We define a map $\mu_{\mathbf{x},u} : \mathcal{APR}_{\mathbf{x},u} \to \{\mathtt{1}, \dots, \#(\mathcal{APR}_{\mathbf{x},u})\} =: A_{\mathbf{x},u}$ analogous to $\Lambda_{\mathbf{x},u}$. The *abelian derived sequence* $\mathcal{E}_u(\mathbf{x})$ is the corresponding infinite word

$$\mu_{\mathbf{x},u}(m_0)\mu_{\mathbf{x},u}(m_1)\mu_{\mathbf{x},u}(m_2)\cdots$$

over $A_{\mathbf{x},u}$ where the sequence $m_0m_1m_2\cdots \in \mathcal{APR}_{\mathbf{x},u}^{\omega}$ is the one computed in the previous lemma. The inverse map $\mu_{\mathbf{x},u}^{-1}$ defines a morphism $\theta_{\mathbf{x},u}$ from $A_{\mathbf{x},u}^*$ to $\mathcal{APR}_{\mathbf{x},u}^*$

Observe that $\mathcal{E}_u(\mathbf{x})$ is uniformly recurrent. Indeed, if $a_1\cdots a_n$ is a factor occurring in $\mathcal{E}_u(\mathbf{x})$, it comes from a factor $m_1\cdots m_n \in \mathcal{APR}_{\mathbf{x},u}^*$ such that $m_1\cdots m_n v$ occurs in $\mathbf{x}$ for some $v \sim_{ab} u$ and $\mu_{\mathbf{x},u}(m_1)\cdots\mu_{\mathbf{x},u}(m_n) = a_1\cdots a_n$. Since $\mathbf{x}$ is uniformly recurrent, the factor $m_1\cdots m_n v$ occurs infinitely often with bounded gaps in $\mathbf{x}$.

Example 3.27. Consider the Thue–Morse word $\mathbf{t}$ introduced in Example 1.11 and its prefix $u = \mathtt{011}$ of length 3. By drawing a vertical line, we mark the occurrences of u in the following prefix of $\mathbf{t}$:

$$|\mathtt{011010}|\mathtt{011001}|\mathtt{01101001}|\mathtt{0110}|\mathtt{011010}|\mathtt{011001}|\mathtt{0110}|\mathtt{01101001}|\mathtt{011010}|\mathtt{011}.$$

We can easily check that the set of return words to u is

$$\mathcal{R}_{\mathbf{t},u} = \{\mathtt{011010}, \mathtt{011001}, \mathtt{01101001}, \mathtt{0110}\}$$

where the words are written in the order of their first occurrence in $\mathbf{t}$. Hence, the set $R_{\mathbf{t},u}$ is $\{\mathtt{1},\mathtt{2},\mathtt{3},\mathtt{4}\}$ and the map $\Lambda_{\mathbf{t},u} : \mathcal{R}_{\mathbf{t},u} \to R_{\mathbf{t},u}$ is defined by

$$\Lambda_{\mathbf{t},u} : \begin{cases} \mathtt{011010} & \mapsto \mathtt{1} \\ \mathtt{011001} & \mapsto \mathtt{2} \\ \mathtt{01101001} & \mapsto \mathtt{3} \\ \mathtt{0110} & \mapsto \mathtt{4}. \end{cases}$$

The derived sequence $\mathcal{D}_u(\mathbf{t})$ is given by

$$\mathcal{D}_u(\mathbf{t}) = \mathtt{1234124312343124123412431241234312341243123431241 2}\cdots$$

Observe that this sequence is the fixed point of the morphism given by

$$\begin{cases} \mathtt{1} \mapsto \mathtt{12} \\ \mathtt{2} \mapsto \mathtt{34} \\ \mathtt{3} \mapsto \mathtt{124} \\ \mathtt{4} \mapsto \mathtt{3}. \end{cases}$$

Similarly in the abelian context, we draw a line to mark the occurrences of of factors abelian equivalent to $u = \mathtt{011}$ in the prefix or $\mathbf{t}$:

$$|\mathtt{0}|\mathtt{1}|\mathtt{1010}|\mathtt{0}|\mathtt{1100}|\mathtt{1}|\mathtt{0}|\mathtt{1}|\mathtt{10100}|\mathtt{1}|\mathtt{0}|\mathtt{110}|\mathtt{0}|\mathtt{1}|\mathtt{1010}|\mathtt{0}|\mathtt{1100}|\mathtt{1}|\mathtt{0}|\mathtt{110}|\mathtt{0}|\mathtt{1}|\mathtt{10100}|\mathtt{1}|\mathtt{0}|\mathtt{1}|\mathtt{1010}|\mathtt{011}.$$

The abelian derived sequence over $A_{\mathbf{t},u} = \{\mathtt{1}, \ldots, \mathtt{6}\}$ is then

$$\mathcal{E}_u(\mathbf{t}) = \mathtt{1231421252161231421612521231421252161252123142161 2}\cdots$$

where the set of abelian returns to u in order of occurrence in $\mathbf{t}$ is given by

$$\mathcal{APR}_{\mathbf{t},u} = \{\mathtt{0},\ \mathtt{1},\ \mathtt{1010},\ \mathtt{1100},\ \mathtt{10100},\ \mathtt{110}\}.$$

Note that, since $\mathtt{0}, \mathtt{1} \in \mathcal{APR}_{\mathbf{t},u}$, there are infinitely many factorizations of $\mathbf{t}$ in terms of elements belonging to $\mathcal{APR}_{\mathbf{t},u}$. In other words, the set $\mathcal{APR}_{\mathbf{t},u}$ is not a code.

Proposition 3.28.

Let u be a prefix of a uniformly recurrent word $\mathbf{x}$. There exists a morphism h_u from $R_{\mathbf{x},u}$ to $A^*_{\mathbf{x},u}$ such that $h_u(\mathcal{D}_u(\mathbf{x})) = \mathcal{E}_u(\mathbf{x})$.

Proof. Each return word m occurring in $\mathbf{x}$ is followed by u. Consider the procedure of Lemma 3.25 applied to mu. It will define the image by h_u of $\Lambda_{\mathbf{x},u}(m)$. Indeed, one has to take into account a factor u appended to m because some suffix of m and a prefix of u can give a word $v \sim_{ab} u$ leading to some abelian return in the decomposition of m. More precisely, we consider all the occurrences $0 = i_1 < \cdots < i_t = |m|$ of factors abelian equivalent to u in $w = mu$. Then

$$h_u(\Lambda_{\mathbf{x},u}(m)) := \mu_{\mathbf{x},u}(w[i_1, i_2 - 1]) \cdots \mu_{\mathbf{x},u}(w[i_{t-1}, i_t - 1]).$$

$\square$

Example 3.29 (Example 3.27 continued)**.** There exists a morphism h_u from $R_{\mathbf{t},u}$ to $A^*_{\mathbf{t},u}$ such that $h_u(\mathcal{D}_u(\mathbf{t})) = \mathcal{E}_u(\mathbf{t})$. Take

$$h_u(\mathtt{1}) = \mathtt{123},\ h_u(\mathtt{2}) = \mathtt{142},\ h_u(\mathtt{3}) = \mathtt{1252},\ h_u(\mathtt{4}) = \mathtt{16}.$$

Let us explain how to get $h_u(\mathtt{2})$. We have the following factorization where the vertical bars indicate the occurrence of a factor abelian equivalent to u:

$$\Lambda^{-1}_{\mathbf{t},u}(\mathtt{2})\, u = (|\mathtt{0}|\mathtt{1100}|\mathtt{1})\, \mathtt{011}.$$

Definition 3.30. A map $h : A^\omega \to B^\omega$ is a *t-block morphism*, if there exists some map $f : A^t \to B^*$ such that, for all $\mathbf{w} \in A^\omega$,

$$h(\mathbf{w}) = f(\mathbf{w}[0, t-1]) f(\mathbf{w}[1, t]) f(\mathbf{w}[2, t+1]) \cdots .$$

By abuse of notation, the second map f will also be denoted by h.

Proposition 3.31.

Let u be a prefix of a uniformly recurrent word $\mathbf{x}$. Let v be a prefix of $\mathbf{y} = \mathcal{E}_u(\mathbf{x})$. There exist $t \le |u| - 1$ and a t-block morphism $h_{u,v} : (A_{\mathbf{y},v})^t \to A^*_{\mathbf{x},u}$ such that

$$h_{u,v}(\mathcal{E}_v(\mathcal{E}_u(\mathbf{x}))) = \mathcal{E}_u(\mathbf{x}).$$

Proof. Note that any element $\theta_{\mathbf{x},u}(\theta_{\mathbf{y},v}(a))$ with $a \in A_{\mathbf{y},v}$ is a concatenation of abelian returns to u. Now consider a factor $a_0 a_1 \cdots a_{t-1}$ occurring in $\mathcal{E}_v(\mathcal{E}_u(\mathbf{x}))$. We have to determine the unique factorization of $\theta_{\mathbf{x},u}(\theta_{\mathbf{y},v}(a_0))$ with abelian returns to u given by Lemma 3.25. This one is completely determined when one knows the $|u| - 1$ symbols occurring next. Without that extra knowledge we cannot uniquely determine the factorization for the last $|u| - 1$ symbols possibly occurring in $\theta_{\mathbf{x},u}(\theta_{\mathbf{y},v}(a_0))$. This is the reason to consider the suffix $a_1 \cdots a_{t-1}$ in such a way that $\theta_{\mathbf{x},u}(\theta_{\mathbf{y},v}(a_1)) \cdots \theta_{\mathbf{x},u}(\theta_{\mathbf{y},v}(a_{t-1}))$ has length at least $|u| - 1$. One takes t large enough to ensure this property for any initial symbol $a_0 \in A_{\mathbf{y},v}$. More precisely, consider all the occurrences $0 = i_1 < \cdots < i_s$ of factors abelian equivalent to u in $w = \theta_{\mathbf{x},u}(\theta_{\mathbf{y},v}(a_0)) \cdots \theta_{\mathbf{x},u}(\theta_{\mathbf{y},v}(a_{t-1}))$. Let r be the largest integer such that $i_r < |\theta_{\mathbf{x},u}(\theta_{\mathbf{y},v}(a_0))|$. Then

$$h_{u,v}(a_0 \cdots a_{t-1}) := \mu_{\mathbf{x},u}(w[i_1, i_2 - 1]) \cdots \mu_{\mathbf{x},u}(w[i_r, i_{r+1} - 1]).$$

Note that the above definition is only meaningful if $a_0 \cdots a_{t-1}$ is a factor of $\mathcal{E}_v(\mathcal{E}_u(\mathbf{x}))$. Since this is the only relevant situation, in any other case, the image of $h_{u,v}$ is set to ε. ☐

Observe that if we iterate the process, since the composition of a t-block morphism and an s-block morphism is an (st)-block morphism, then there exists an r-block morphism h such that $h(\mathcal{E}_{u_k}(\cdots(\mathcal{E}_{u_2}(\mathcal{E}_{u_1}(\mathbf{x})))\cdots)) = \mathcal{E}_{u_1}(\mathbf{x})$ where prefixes $u_1, \ldots, u_k$ are considered accordingly.

Example 3.32 (Example 3.27 continued)**.** We can iterate the process of computing the abelian derived sequence, for instance by taking each time the corresponding prefix of length 3:

$$\begin{aligned}
&\mathcal{E}_{\mathtt{123}}(\mathcal{E}_{\mathtt{011}}(\mathbf{t}))\\
&\qquad = \mathtt{121314151213151412131415141213151213141512131514 12}\cdots,\\
&\mathcal{E}_{\mathtt{121}}(\mathcal{E}_{\mathtt{123}}(\mathcal{E}_{\mathtt{011}}(\mathbf{t})))\\
&\qquad = \mathtt{123412431234312412341243124123431234124312343124 12}\cdots,\\
&\mathcal{E}_{\mathtt{123}}(\mathcal{E}_{\mathtt{121}}(\mathcal{E}_{\mathtt{123}}(\mathcal{E}_{\mathtt{011}}(\mathbf{t}))))\\
&\qquad = \mathtt{123414321234321412341432141234321234143212343214 12}\cdots.
\end{aligned}$$

Let us illustrate the previous result. Take again $u = \mathtt{011}$, $\mathbf{y} = \mathcal{E}_u(\mathbf{t})$, $v = \mathtt{123}$. We have

$$\mathcal{APR}_{\mathbf{y},v} = \{\mathtt{1},\ \mathtt{23142125216},\ \mathtt{23142161252},\ \mathtt{231421252161252},\ \mathtt{2314216}\}.$$

Observe that $\theta_{\mathbf{t},u}(\theta_{\mathbf{y},v}(\mathtt{1})) = \theta_{\mathbf{t},u}(\mathtt{1}) = \mathtt{0}$ and, for all $a \in \{\mathtt{2}, \ldots, \mathtt{5}\}$, $\theta_{\mathbf{y},v}(a)$ has a prefix $\mathtt{23}$, so $\theta_{\mathbf{t},u}(\theta_{\mathbf{y},v}(a))$ has prefix $\mathtt{110} \sim_{ab} u$. Let us assume that $h_{u,v}$ is a 3-block morphism. We define $h_{u,v}(\mathtt{1}ab) = \mathtt{1}$, for all $a, b \in A_{\mathbf{y},v}$ and $a \neq \mathtt{1}$. We get $h_{u,v}(\mathtt{213}) = \mathtt{23142125216}$ because, if vertical bars denote occurrences of a factor abelian equivalent to u, we get the following factorization:

$$\theta_{\mathbf{t},u}(\theta_{\mathbf{y},v}(2))\,\theta_{\mathbf{t},u}(\theta_{\mathbf{y},v}(13))$$
$$= (|1|1010|0|1100|1|0|1|10100|1|0|110)\,|0110100110010110011010 01.$$

Proposition 3.33.

Let ζ be a primitive substitution and u be a prefix of its fixed point $\mathbf{x} = \zeta(\mathbf{x}) \in A^\omega$. There exists a 2-block morphism $\zeta_u : A^*_{\mathbf{x},u} \to A^*_{\mathbf{x},u}$ such that $\mathcal{E}_u(\mathbf{x})$ is fixed point of ζ_u and

$$\theta_{\mathbf{x},u}(\zeta_u(\mathcal{E}_u(\mathbf{x}))) = \zeta(\theta_{\mathbf{x},u}(\mathcal{E}_u(\mathbf{x}))).$$

Proof. We may replace ζ by a convenient power of ζ in such a way that, for all $a \in A$, $\zeta(a)$ contains an occurrence of a factor abelian equivalent to u. For all $a, b \in A_{\mathbf{x},u}$, we consider all the occurrences $i_1 < \cdots < i_t$ of a factor abelian equivalent to u occurring in $w = \zeta(\theta_{\mathbf{x},u}(ab))$. With our choice of ζ, at least one of these i_j belongs to $[0, |\zeta(\theta_{\mathbf{x},u}(a))| - 1]$ (resp. $[|\zeta(\theta_{\mathbf{x},u}(a))|, |w| - 1]$). Let r be the largest integer such that $i_r < |\zeta(\theta_{\mathbf{x},u}(a))|$. We define

$$\zeta_u(ab) = \mu_{\mathbf{x},u}(w[i_1, i_2 - 1]) \cdots \mu_{\mathbf{x},u}(w[i_r, i_{r+1} - 1]).$$

□

Corollary 3.34.

Let ζ be a primitive substitution and u be a prefix of its fixed point $\mathbf{x} = \zeta(\mathbf{x}) \in A^\omega$. The sequence $\mathcal{E}_u(\mathbf{x})$ is primitive substitutive, i.e., there exists a primitive morphism $\varphi_u : B \to B^*$ and a coding $\phi : B \to A_{\mathbf{x},u}$ such that $\mathcal{E}_u(\mathbf{x}) = \phi(\varphi_u^\omega(b))$ for some $b \in B$.

Proof. We may replace ζ by a convenient power of ζ in such a way that, for all $a \in A$, $\zeta(a)$ contains occurrences of two factors abelian equivalent to u. Consider the alphabet

$$B = \{(a, b) \mid a, b \in A_{\mathbf{x},u} \wedge ab \text{ is a factor of } \mathcal{E}_u(\mathbf{x})\}.$$

For all $a, b \in A_{\mathbf{x},u}$ such that $(a, b) \in B$, consider all the occurrences $i_1 < \cdots < i_t$ of a factor abelian equivalent to u occurring in $w = \theta_{\mathbf{x},u}(ab)$. Let r be the smallest integer such that $i_r \geq |\theta_{\mathbf{x},u}(a)|$. Note that $r \geq 3$. We define

$$\varphi_u((a, b)) = (\mu_{\mathbf{x},u}(w[i_1, i_2 - 1]), \mu_{\mathbf{x},u}(w[i_2, i_3 - 1])) \cdots$$
$$(\mu_{\mathbf{x},u}(w[i_{r-1}, i_r - 1]), \mu_{\mathbf{x},u}(w[i_r, i_{r+1} - 1])).$$

Let $e_0 e_1$ be the prefix of length 2 of $\mathcal{E}_u(\mathbf{x})$. We have

$$\mathcal{E}_u(\mathbf{x}) = \phi(\varphi_u^\omega((e_0, e_1)))$$

where $\phi : B \to A_{\mathbf{x},u}$ is the coding that maps $(a, b) \in B$ to a.

Observe that, for all $(a,b) \in B$, $|\varphi_u^n((a,b))| \geq 2^n$. Let us show that φ_u is primitive. Since $\mathcal{E}_u(\mathbf{x})$ is uniformly recurrent, there exists K such that any factor of length K of $\mathcal{E}_u(\mathbf{x})$ contains all elements in $\{cd \mid (c,d) \in B\}$. Therefore any factor of length K of $\varphi_u^\omega((e_0,e_1))$ contains all the elements of B. Take N such that $2^N \geq K$. Then, for all $(a,b),(c,d) \in B$, $\varphi_u^N((a,b))$ contains (c,d) which means that φ_u is primitive. $\square$

Example 3.35 (Example 3.27 continued)**.** Take again $u = \mathtt{011}$ and set ζ to be the morphism σ^3 defined by $\mathtt{0} \mapsto \mathtt{01101001}, \mathtt{1} \mapsto \mathtt{10010110}$ that generates $\mathbf{t}$. We have

$$\zeta(\theta_{\mathbf{t},u}(\mathtt{12})) = (|\mathtt{0}|\mathtt{1}|\mathtt{1010}|\mathtt{0}|\mathtt{1})\mathtt{100}|\mathtt{1}|\mathtt{0}|\mathtt{110} \text{ and } \zeta_u(\mathtt{12}) = \mathtt{12314}$$

$$\zeta(\theta_{\mathbf{t},u}(\mathtt{23})) = (\mathtt{100}|\mathtt{1}|\mathtt{0}|\mathtt{1}|\mathtt{10})\mathtt{100}|\mathtt{1}|\mathtt{0}|\mathtt{110}\cdots \text{ and } \zeta_u(\mathtt{23}) = \mathtt{2125}$$

$$\zeta(\theta_{\mathbf{t},u}(\mathtt{31})) = (\mathtt{100}|\mathtt{1}|\mathtt{0}|\mathtt{110}|\mathtt{0}|\mathtt{1}|\mathtt{1010}|\mathtt{0}|\mathtt{1100}|\mathtt{1}|\mathtt{0}|\mathtt{110}|\mathtt{0}|\mathtt{1}|\mathtt{10100}|\mathtt{1})|\mathtt{0}|\mathtt{1}|\mathtt{101001}$$

$$\text{and } \zeta_u(\mathtt{21}) = \mathtt{216123142161252}.$$

Using the above corollary, we get

$$\varphi_u(1,2) = (1,2)(2,3)(3,1)(1,4)(4,2),\ \varphi_u(2,3) = (2,1)(1,2)(2,5)(5,2),\ldots.$$

3.4.1 There are infinitely many abelian derivatives of the Thue–Morse word

Proposition 3.36.

For the Thue–Morse word $\mathbf{t}$, the set $\{\mathcal{E}_u(\mathbf{t}) \mid u \in \mathrm{Pref}(\mathbf{t})\}$ is infinite.

Proof. It is sufficient to show that the set $\{\mathcal{E}_u(\mathbf{t}) \mid u \in \mathrm{Pref}(\mathbf{t}) : |u| \equiv 1 \pmod 2\}$ is infinite. Proceed by contradiction and suppose that the set $\{\mathcal{E}_u(\mathbf{t}) \mid u \in \mathrm{Pref}(\mathbf{t}) : |u| \equiv 1 \pmod 2\}$ is finite. Then there exist u and v distinct prefixes of odd length of the Thue–Morse word $\mathbf{t}$ such that $\mathcal{E}_u(\mathbf{t}) = \mathcal{E}_v(\mathbf{t})$. Since $\mathcal{APR}_{\mathbf{t}}$ is finite, we can moreover assume that $\theta_{\mathbf{t},u} = \theta_{\mathbf{t},v}$. Indeed, infinitely many sequences of the kind $\mathcal{E}_u(\mathbf{t})$ are equal and thus defined on the same alphabet $A_{\mathbf{t},u}$. For all such sequences, there are finitely many morphisms of the kind $\theta_{\mathbf{t},u}$ associating with each element of $A_{\mathbf{t},u}$ an element of the finite set $\mathcal{APR}_{\mathbf{t}}$. So we can impose the extra condition on $\theta_{\mathbf{t},u}$. Let

$$I(w) := \{i \in \mathbb{N} \mid \mathbf{t}[i, i+|w|-1] \sim_{ab} w\}$$

denote the set of occurrences of factors of $\mathbf{t}$ abelian equivalent to a word w. We have $I(u) = I(v)$ as $\theta_{\mathbf{t},u} = \theta_{\mathbf{t},v}$. Without loss of generality, we may suppose that $|u| = 2k+1$, $|v| = 2\ell+1$ with $k < \ell$. We have $\Psi(u) = (k, k+1)$ or $\Psi(u) = (k+1, k)$ and $\Psi(v) = (\ell, \ell+1)$ or $\Psi(v) = (\ell+1, \ell)$. Let a_u (resp. a_v) denote the letter having $k+1$ (resp. $\ell+1$) occurrences in the prefix u (resp. v). Note that a_u and a_v are respectively the last letters of u and v.

For any odd position j, recalling that $\mathbf{t}[2m, 2m+1] \in \{\mathtt{10}, \mathtt{01}\}$, we have

$$\mathbf{t}_j = a_u \Leftrightarrow \mathbf{t}[j, j+|u|-1] \sim_{ab} u \Leftrightarrow \mathbf{t}[j, j+|v|-1] \sim_{ab} v \Leftrightarrow \mathbf{t}_j = a_v$$

where the central equivalence comes from the fact that $I(u) = I(v)$. As there exists at least one such j, we have $a_u = a_v =: a$.

For any even position j, we have

$$\mathbf{t}_{j+|u|-1} = a \Leftrightarrow j \in I(u) \Leftrightarrow j \in I(v) \Leftrightarrow \mathbf{t}_{j+|v|-1} = a$$

since $I(u) = I(v)$. Using this observation, we can show by induction that

$$\mathbf{t}_{|u|-1+n(|v|-|u|)} = a$$

for all $n \in \mathbb{N}$. In other words, there exists a constant infinite arithmetical subsequence in $\mathbf{t}$, which is a contradiction, since it is well-known that the Thue-Morse word does not contain any such subsequence. Indeed, for $n = 0$, it is clear that the last letter of u is a. Suppose now that the result holds true for $n \geq 0$. We have $\mathbf{t}_{|u|-1+n(|v|-|u|)} = a$. Since $|u|, |v|$ are odd, $n(|v| - |u|)$ is an even number and belongs to $I(u) = I(v)$. Therefore $\mathbf{t}_{|v|-1+n(|v|-|u|)} = a$ and $|v| - 1 + n(|v| - |u|) = |u| - 1 + (n + 1)(|v| - |u|)$. □

Remark 3.37. Using the same notation as in the previous proof, we show that the set $\{\mathcal{E}_u(\mathbf{t}) \mid u \in \mathrm{Pref}(\mathbf{t}) : |u| \equiv 0 \pmod 2\}$ is infinite. Proceed by contradiction. Then there exist u and v distinct prefixes of even length of $\mathbf{t}$ such that $\mathcal{E}_u(\mathbf{t}) = \mathcal{E}_v(\mathbf{t})$ and $\theta_{\mathbf{t},u} = \theta_{\mathbf{t},v}$. Hence $I(u) = I(v)$. Note that $2\mathbb{N} \subseteq I(u) = I(v)$. Suppose that $|u| = 2k$, $|v| = 2\ell$, with $k < \ell$. Since a prefix of even length has a Parikh vector of the kind (r, r), $2i+1$ is in $I(u)$ if and only if $\mathbf{t}_i = \mathbf{t}_{i+k}$. Similarly, $2i + 1$ is in $I(v)$ if and only if $\mathbf{t}_i = \mathbf{t}_{i+\ell}$. From $I(u) \setminus 2\mathbb{N} = I(v) \setminus 2\mathbb{N}$, we deduce that, for all $i \in \mathbb{N}$, $\mathbf{t}_i = \mathbf{t}_{i+k}$ implies $\mathbf{t}_i = \mathbf{t}_{i+\ell}$ and conversely. This leads to the contradiction that $\mathbf{t}$ is ultimately periodic of period $\ell - k$. Indeed, suppose to the contrary that for some i, $\mathbf{t}_{i+k} \neq \mathbf{t}_{i+\ell}$. In this case, either $\mathbf{t}_{i+k}$ or $\mathbf{t}_{i+\ell}$ is equal to $\mathbf{t}_i$. From our last deduction, we get that all three letters $\mathbf{t}_i$, $\mathbf{t}_{i+k}$, $\mathbf{t}_{i+\ell}$ are equal.

Remark 3.38. Using the same notation as in the previous remark, there exist no prefixes u, v of $\mathbf{t}$ such that $|u|$ is even, $|v|$ is odd and $I(u) = I(v)$. (The symmetric case can be treated in the same way.) Assume that $|u| = 2k$, $|v| = 2\ell + 1$ for some positive integers $k \neq \ell$. We get $\Psi(u) = (k, k)$ and $\Psi(v) = (\ell, \ell + 1)$ or $\Psi(v) = (\ell + 1, \ell)$. Let a denote the letter that has $\ell + 1$ occurrences in v. As $v \in \mathrm{Pref}(\mathbf{t})$, $\mathbf{t}_{|v|-1} = a$. Note that, for all even positions j, if j is in $I(v)$, then $\mathbf{t}_{j+|v|-1} = a$. Moreover, for all even j, we have $j \in I(u)$. Since $I(u) = I(v)$, we get $2\mathbb{N} \subseteq I(v)$ and thus $\mathbf{t}_{j+|v|-1} = a$ for all even j. Therefore, for all even $j \geq |v| - 1$, we have $\mathbf{t}_j = a$ and also $\mathbf{t}_{j+1} = \mathtt{1} - a$ since $\mathbf{t}$ is made up of blocks $\mathtt{01}$ or $\mathtt{10}$. This means that the Thue–Morse word is ultimately periodic of period 2 which is a contradiction.

3.5 Related work and some open questions

Masáková and Pelantová complete our work about Sturmian words [MP13]. While we only describe the set of abelian returns to all prefixes for a particular Sturmian word, namely the Fibonacci word, they give the set of abelian returns to all prefixes for any characteristic Sturmian word, using the continued fraction expansion of α [MP13, Proposition 18]. We let $[0, a_1, a_2, \ldots]$ denote the continued fraction expansion of a number a in $(0, 1)$:

$$a = 0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \cdots}}.$$

Proposition 3.39. [MP13]

Let $\alpha = [0, a_1, a_2, \ldots]$ be an irrational in $(0, 1)$. For the characteristic Sturmian word $\mathbf{c}_\alpha$, we have

$$\mathcal{APR}_{\mathbf{c}_\alpha} = \begin{cases} \{\mathtt{0}, \mathtt{01}, \mathtt{1}, \mathtt{10}, \mathtt{110}, \ldots, \mathtt{1}^{a_1}\mathtt{0}\} & \text{if } \alpha < \frac{1}{2} \\ \{\mathtt{1}, \mathtt{10}, \mathtt{0}, \mathtt{01}, \mathtt{001}, \ldots, \mathtt{0}^{a_2+1}\mathtt{1}\} & \text{otherwise.} \end{cases}$$

Let $\left(\frac{p_n}{q_n}\right)_{n\geq 0}$ denote the sequence of the convergents associated with α:

$$\frac{p_n}{q_n} = [0, a_1, \ldots, a_n].$$

The numerators p_n and the denominators q_n satisfy the recurrence relations

$$p_k = a_k p_{k-1} + p_{k-2} \text{ and } q_k = a_k q_{k-1} + q_{k-2}$$

for every $k \geq 1$ with initial values $p_0 = a_0 = 0$, $q_0 = 1$, $p_{-1} = 1$ and $q_{-1} = 0$. We set $\delta_{k,s} := |(s-1)(p_k - \alpha q_k) + p_{k-1} - \alpha q_{k-1}|$ for $k \geq 0$, $1 \leq s \leq q_{k+1}$. Then one has $\delta_{k,s} < \delta_{k',s'}$ if and only if ks is greater than $k's'$ for the lexicographic order. Masáková and Pelantová determine the cardinality of the set $\mathcal{APR}_\mathbf{x}$ for any Sturmian word $\mathbf{x}$ [MP13, Theorem 16].

Theorem 3.40. [MP13]

Let $\alpha, \rho \in (0, 1)$, $\alpha = [0, a_1, a_2, \ldots]$ irrational with $a_1 \geq 2$. Let $\mathbf{x}$ be a Sturmian word with slope α and intercept ρ.

- For $\rho \in (\alpha, 1 - \alpha)$, $\#\mathcal{APR}_\mathbf{x} \in \{a_1 + 3, a_1 + 4\}$.
- For $\rho \notin (\alpha, 1-\alpha)$, let $k, s \in \mathbb{N}$, with $1 \leq s \leq a_k+1$, be minimal in lexicographic order such that $\min\{\rho, 1 - \rho\} \geq \delta_{k,s}$. Then $\#\mathcal{APR}_\mathbf{x} = 2 + a_1 + \cdots + a_k + s$.

Also, Masáková and Pelantová slightly extend Puzynina and Zamboni's work [PZ13] and give an alternative proof of their characterization of Sturmian words. Puzynina and Zamboni show that an aperiodic recurrent infinite word $\mathbf{x}$ is Sturmian if and only if every factor of $\mathbf{x}$ has two or three abelian return words, i.e., $\#\mathcal{APR}_{\mathbf{x},u} \in \{2, 3\}$ for all $u \in \mathrm{Fac}(\mathbf{x})$. If a factor has three abelian return words R_1, R_2, R_3, then $|R_1| + |R_2| = |R_3|$. Masáková and Pelantová show that in fact $R_3 = R_1 R_2$ [MP13].

In Section 1.5, we presented Sturmian words through different equivalent definitions. In particular, Sturmian words are a particular class of rotation words (Proposition 1.63). Rampersad et al. study abelian returns in a more general class of rotation words [RRS14]. In addition to the angle α and the intercept ρ, they consider a third parameter β and a partition the unit circle into intervals $I_0 = [0, 1 - \beta)$ and $I_1 = [1 - \beta, 1)$. So $r(\alpha, \beta, \rho)$ is the coding of the trajectory of the point ρ under a rotation of angle α with respect to the intervals I_0, I_1. Sturmian words in this setting are words $r(\alpha, \alpha, \rho)$ with α irrational. The main contribution of Rampersad et al. is a characterization analogous to Theorem 3.18.

Theorem 3.41. [RRS14]

Let α be irrational and $m \geq 1$ be an integer. Let $\mathbf{r} = r(\alpha, \{m\alpha\}, \rho)$ be a rotation word. The set $\mathcal{APR}_{\mathbf{r}}$ is finite if and only if $\rho \notin \{\{-i\alpha\} \mid 0 \leq i < m\}$.

In addition, Rampersad et al. [RRS14] prove that one direction of Theorem 3.1 is a consequence of the *three gap theorem*. Their proof uses our number-theoretic approach, in particular Equation (3.3).

Theorem 3.42. Three gap theorem [AB98]

Let ρ be a real number, $\alpha \in (0,1)$ be irrational and let I be a proper subinterval of $(0,1)$. The gaps between the successive integers j such that $\{j\alpha + \rho\} \in I$ take at most three values, one being the sum of the other two.

In Subsection 3.3.2 and Subsection 3.3.3, we used the bijection $v \mapsto I_v$ that maps a word to an interval. Fici et al. [FLL$^+$13] use a similar Sturmian bijection to study abelian repetitions in Sturmian words. The authors show that the Sturmian bijection preserves abelian properties of factors. Hence they are able to apply number-theoretic techniques to obtain the following result.

Theorem 3.43. [FLL$^+$13]

Let $\mathbf{x}$ be a Sturmian word. For any integer $m > 1$, let k_m denote the maximal exponent of an abelian repetition of period m in $\mathbf{x}$. Then

$$\limsup_{m \to \infty} \frac{k_m}{m} \geq \sqrt{5}$$

and the equality holds for Sturmian words with slope $1/\phi^2$.

Hence, with the same notation as in the previous theorem, we have $\limsup \frac{k_m}{m} = \sqrt{5}$ for the Fibonacci word. We conclude this chapter with two open questions that derives from the previous theorem and that are given in Fici et al. [FLL$^+$13].

- Is it possible to find the exact value of $\limsup \frac{k_m}{m}$ for other Sturmian words s_α with slope α different from $1/\phi^2$?
- Is it possible to give the exact value of this superior limit when α is an algebraic number of degree 2?

Part II

Covering problems in graphs

Chapter 4

Graphs

This chapter contains all basic notions used in Part 2. First, we recall some usual definitions and results about graphs. Secondly, some product operations on graphs are defined. Next, we define well-known classes of graphs. We present then colourings of the vertices of a graph. Finally, we consider the notion of covering problems. We consider in particular the problem of finding an identifying code of minimal size and a covering problem with multiplicity conditions.

Contents

4.1 Basic definitions about graphs

We briefly introduce the basic terminology used in theory of graphs. More details are given in [Die10]. An *undirected graph* G is given by an ordered pair (V, E) where $E \subseteq V \times V$ is a symmetric binary relation on a set V. We call the elements of V *vertices* and the elements of E *edges*. When the sets V and E are not given explicitly, we denote by $V(G)$ and $E(G)$ the set of vertices and edges of the graph G. Since E is a symmetric relation, an edge between a vertex u and a vertex v is written as a non-ordered pair $\{u, v\}$. If the context is clear, we write $\{u, v\}$ simply as uv or vu. An edge of the form uu is a *loop*. A graph without any loop is said to be *simple*.

Remark 4.1. In this thesis, we only consider graphs that are undirected and simple.

The usual way to picture a graph is by drawing a dot or a small circle for each vertex and joining two of these dots by a line if the corresponding two vertices form an edge. For instance, Figure 4.1 depicts an undirected graph with vertex set $V = \{v_0, \ldots, v_6\}$ and edge set

$$E = \{v_0v_1, v_0v_4, v_1v_2, v_1v_4, v_2v_3, v_2v_5, v_2v_6, v_4v_5\}.$$

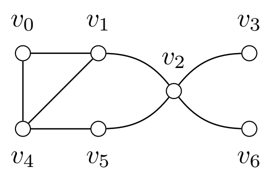


Figure 4.1: A graph with 7 vertices.

Two vertices u and v are *adjacent* if uv is an edge. Two adjacent vertices are called *neighbours*. If all the vertices of G are pairwise adjacent, then the graph is called *complete*. We denote by K_n the complete graph on n vertices.

The set of all neighbours of u is the *open neighbourhood* denoted by $N(u)$:

$$N(u) = \{v \in V \mid uv \in E\}.$$

The *closed neighbourhood* of u is $N[u] = N(u) \cup \{u\}$, i.e., u and all its neighbours. The *degree* of a vertex is the number of its neighbours. A graph is *regular* if all vertices have the same degree.

Example 4.2. Consider the graph represented in Figure 4.1. The vertex v_0 has degree $\deg(v_0) = 2$. The open neighbourhood of v_1 is $N(v_1) = \{v_0, v_2, v_4\}$ and the closed neighbourhood of v_3 is $N[v_3] = \{v_2, v_3\}$.

A set of vertices that are not adjacent is called *independent*. Similarly, edges without common vertices are *independent*. A set M of independent edges in a graph G is called a *matching*. We say the matching is *perfect* if any vertex of the graph is an endpoint of an edge of M.

Example 4.3. Let G be the graph depicted in Figure 4.1. The set $\{v_0, v_5, v_6\}$ of vertices is independent. The set $\{v_0v_1, v_2v_3, v_4v_5\}$ of edges is a matching of the graph. Moreover, one can check that the graph does not have any perfect matching.

Consider now the graph depicted in Figure 4.2(a). It has 3 perfect matchings, one of them is represented in Figure 4.2(b).

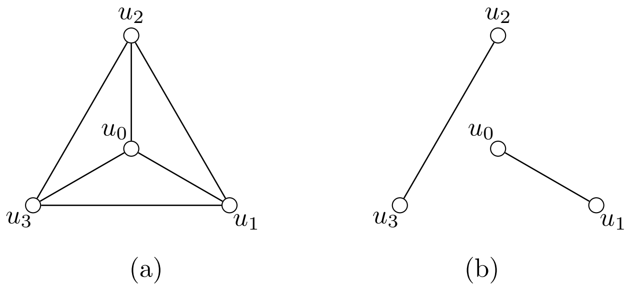


Figure 4.2: A graph (a) and one of its perfect matchings (b).

Let $k \geq 2$ be an integer. A graph $G = (V, E)$ is *k-partite* or *multipartite* if V admits a partition into k sets such that each edge has its ends in different sets. In other words, each set is independent. Instead of 2-partite, we usually say *bipartite*. A k-partite graph for which every two vertices from different sets are adjacent is called *complete*. The complete k-partite graph for which the cardinal of the set are respectively $n_1, \ldots, n_k$ is denoted by $K_{n_1,\ldots,n_k}$. Examples of 3-partite graphs are given in Figure 4.3, the graph on the right is a complete 3-partite graph.

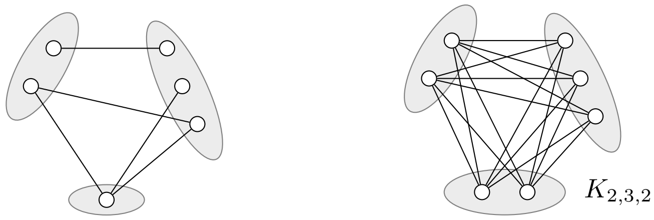


Figure 4.3: Two 3-partite graphs.

Given two vertices u and v of G, the *distance* between u and v is the number of edges in a shortest path in G between u and v. We denote this distance by $d_G(u, v)$, or simply $d(u, v)$ if the context is clear. The *diameter* of a finite graph G, denoted by $\mathrm{diam}(G)$, is the maximum distance between any pair of vertices of the graph. Let r be a positive integer. The *ball* of radius r with center v, denoted by $B_r(v)$ and sometimes called r-ball, is the subset of vertices lying at distance at most r from v:

$$B_r(v) = \{u \in V \mid d(u, v) \leq r\}.$$

In particular, we have $B_1(v) = N[v]$.

Example 4.4. Consider again the graph depicted in Figure 4.1. The distance between v_0 and v_5 is $d(v_0, v_5) = 2$. The ball with radius 2 with center v_0 is $B_2(v_0) = \{v_0, v_1, v_2, v_4, v_5\}$ and the ball of radius 2 with center v_2 is equal to the whole vertex set. The diameter of the graph is 3 and the vertices v_0, v_6 realize the diameter, i.e., $d(v_0, v_6) = 3 = \mathrm{diam}(G)$.

An *isomorphism* $\varphi : G = (V, E) \to G' = (V', E')$ between two graphs G and G' is a bijective application from V to V' that preserves the edges of the graph: uv is an edge of G if and only if $\varphi(u)\varphi(v)$ is an edge of G'. If $G = G'$, φ is called an *automorphism* of G. We denote by $Aut(G)$ the set of all automorphisms of the graph G.

Example 4.5. There are only two automorphisms in the set of automorphisms of the graph given in Figure 4.1, namely the map φ defined by

$$\varphi : \begin{cases} v_i \mapsto v_i & \text{for } i \in \{0, 1, 2, 4, 5\} \\ v_3 \mapsto v_6 \\ v_6 \mapsto v_3 \end{cases}$$

and the identity map.

In some graph, vertices seem to play the same role. Such graphs are called *vertex-transitive*. More formally, a graph is *vertex-transitive* if for any pair of vertices u and v there exists an automorphism sending u to v. A vertex-transitive graph is in particular regular.

Example 4.6. As the graph represented in Figure 4.1 is not regular, it is not vertex-transitive. The graph depicted in Figure 4.2(a) is clearly vertex-transitive and 3-regular.

Observe that not all regular graphs are vertex-transitive.

Example 4.7 (Frucht graph)**.** The graph depicted in Figure 4.4 is a 3-regular graph that only has the identity map as automorphism. Hence, it is not vertex-transitive. The graph is named after Frucht who first described it in 1939 [Fru39].

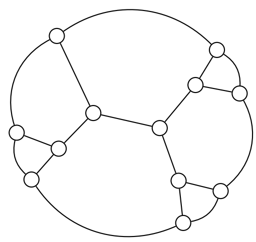


Figure 4.4: The Frucht graph is 3-regular but not vertex-transitive.

4.1.1 A link between graphs and words

We now present a way to obtain a graph from the set of words of a fixed length, with respect to a metric. For more details about words see Section 1.1. Let A be an alphabet with k letters and n be a positive integer. For a metric corresponding to a distance d, we define the graph G with vertex set $V = A^n$ and edge set $E = \{\{u, v\} \in V^2 \mid d(u, v) = 1\}$. In other words, the vertices correspond to words of length n over A and two vertices are adjacent if the corresponding words are at distance 1 with respect to the metric. In the sequel, we mention three well-known metrics, namely the Hamming metric, the Lee metric and the Manhattan metric.

Definition 4.8 (Hamming metric)**.** The *Hamming distance* between two words is the number of different letters appearing at the same index. For instance, the word `01302` is at distance 2 from the word `31102` as the first and the third letters are different. The graphs corresponding to the Hamming metric are called *Hamming graphs*.

Definition 4.9 (Manhattan metric)**.** The *Manhattan distance* between two $u = u_0 \cdots u_{n-1}$ and $v = v_0 \cdots v_{n-1}$ with $u_i, v_i \in A$ is the sum

$$\sum_{i=0}^{n-1} |u_i - v_i|.$$

For example, the words `01302` and `31102` are at distance $|0-3| + |3-1| = 5$.

The Lee metric is similar to the Manhattan metric. They coincide in the case of infinite alphabets.

Definition 4.10 (Lee metric)**.** The *Lee distance* between two words $u = u_0 \cdots u_{n-1}$ and $v = v_0 \cdots v_{n-1}$ with $u_i, v_i \in A$ is the sum

$$\sum_{i=0}^{n-1} \min(|u_i - v_i|, k - |u_i - v_i|)$$

where k is the number of letters in the alphabet A. For instance, the words `01302` and `31102` are at distance 3 since $\min(|0-3|, 4-|0-3|) = 1$ and $\min(|3-1|, 4-|3-1|) = 1$.

4.2 Operations on graphs

Let $G = (V(G), E(G))$ and $H = (V(H), E(H))$ be two graphs. We present first the complement and power of graph, then some products of two graphs.

The *complement* (in terms of simple graphs) of G is the graph $\overline{G}$ with the same vertex set $V(G)$, such that two distinct vertices are adjacent in $\overline{G}$ if and only if they are not adjacent in G. Observe that since we only consider simple graphs, the definition of the complement of a graph implies that the complement is also simple.

Example 4.11. Consider the graph G that is depicted in Figure 4.5. Its vertex set is $V = \{u_0, u_1, u_2, u_3, u_4\}$ and its edge set is $E = \{u_0u_1, u_1u_2, u_2u_3, u_3u_4\}$. The complement $\overline{G}$ of G has edge set

$$E(\overline{G}) = \{u_0u_2, u_0u_3, u_0u_4, u_1u_3, u_1u_4, u_2u_4\}.$$

Figure 4.5: A graph G and its complement $\overline{G}$.

Let r be a positive integer. The *r-power* of G is the graph G^r with vertex set $V(G)$ and such that two vertices are adjacent in G^r if they are at distance at most r in G:

$$E(G^r) = \{uv \in V \times V | d_G(u, v) \leq r\}.$$

Example 4.12. If G is a graph with vertex set $V = \{u_0, u_1, u_2, u_3, u_4\}$ and edge set $E = \{u_0u_1, u_1u_2, u_2u_3, u_3u_4\}$, then G^2 has edge set

$$E(G^2) = E \cup \{u_0u_2, u_1u_3, u_2u_4\}$$

as represented in Figure 4.6.

Figure 4.6: A graph G and its 2-power G^2.

The *Cartesian product* of G and H, denoted by $G \square H$, has vertex set $V(G) \times V(H)$ and two vertices (u_G, u_H) and (v_G, v_H) are adjacent if one of the following holds

- $u_G = v_G$ and $u_H v_H \in E(H)$,
- $u_H = v_H$ and $u_G v_G \in E(G)$.

For example, the Cartesian product $K_3 \square K_4$ of two complete graphs K_3 and K_4 is depicted in Figure 4.7.

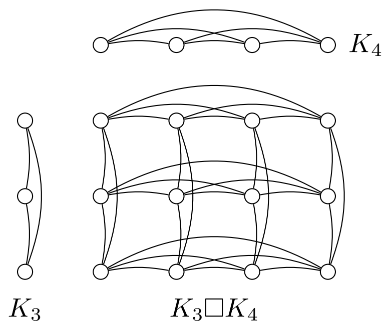


Figure 4.7: Cartesian product of two complete graphs.

The *direct product* of G and H, denoted by $G \times H$, has vertex set $V(G) \times V(H)$ and two vertices (u_G, u_H) and (v_G, v_H) are adjacent if $u_G v_G \in E(G)$ and $u_H v_H \in E(H)$. See Figure 4.8 for an example.

The *lexicographic product* of two graphs G and H, denoted by $G[H]$, has vertex set $V(G) \times V(H)$ and two vertices (u_G, u_H) and (v_G, v_H) are adjacent if one of the following holds

- $u_G v_G \in E(G)$,
- $u_G = v_G$ and $u_H v_H \in E(H)$.

An example is given in Figure 4.9.

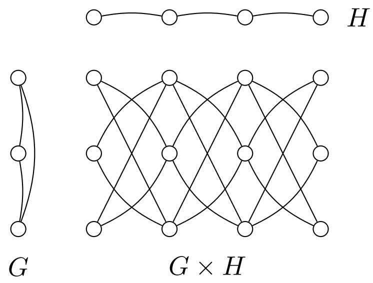


Figure 4.8: Direct product of two graphs.

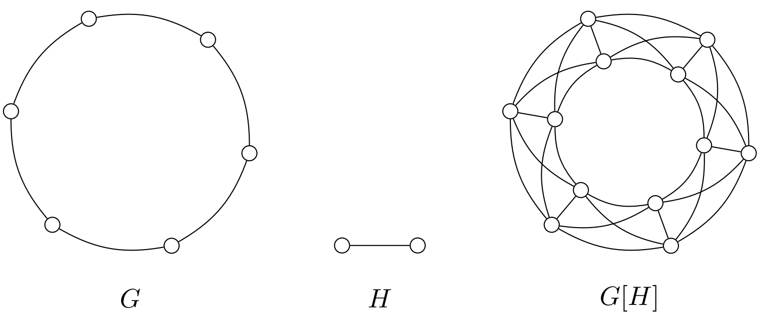


Figure 4.9: Lexicographic product of two graphs.

4.3 Some classes of graphs

In this section, we present different graphs considered in the next two chapters. All graphs in Chapter 5 are finite while the graphs discussed in Chapter 6 are infinite.

4.3.1 Finite paths and cycles

A *path* on n vertices is a graph, denoted by P_n, with vertex-set $\{0, \ldots, n-1\}$ and its edges are the non-ordered pairs $\{i, i+1\}$ for $0 \leq i < n-1$. The Cartesian product of k paths P_n corresponds to the graph obtained from the set of words of length k over a n-letter alphabet with the Manhattan distance.

A *cycle* on n vertices is a graph, denoted by $\mathcal{C}_n$, with vertex-set $\{0 \ldots, n-1\}$. Its edges are the non-ordered pairs $\{i, i+1\}$ for $0 \leq i < n-1$ and $\{0, n-1\}$. Clearly any cycle is vertex-transitive. Observe that the Cartesian product of k cycles $\mathcal{C}_n$ corresponds to the graph obtained from the set of words of length k over a n-letter alphabet with the Lee metric.

The *length* of a finite path or a cycle is the number of its edges. Hence, P_n and $\mathcal{C}_n$ have respective length $n-1$ and n.

4.3.2 Hypercubes

Let $q \geq 3$. The *hypercube of dimension* q is the graph $\mathcal{H}_q$ with the set $\{\mathtt{0}, \mathtt{1}\}^q$ of binary words of length q as vertex set. Two vertices are adjacent if the corresponding words differ on exactly one letter. Hence, the hypercube of dimension q is a Hamming graph. For instance, the hypercubes of dimension 1 to 3 are given in Figure 4.10. Note that we have $\mathcal{H}_1 = P_2$.

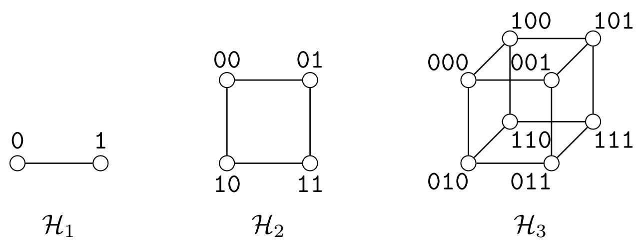


Figure 4.10: The hypercubes of dimension 1 to 3.

Observe that for all $q \geq 1$, the hypercube of dimension q can be recursively constructed by taking two copies of $\mathcal{H}_{q-1}$, then adding edges between the corresponding vertices of each copy (Figure 4.11). In other words, $\mathcal{H}_q = \mathcal{H}_{q-1} \square P_2$. Clearly, $\mathcal{H}_q$ is vertex-transitive with vertex degree q.

4.3.3 Strongly regular graphs

In the next chapter, we often consider strongly regular graphs, that are a particular class of finite graphs.

Definition 4.13. A *strongly regular graph* $\mathrm{srg}(n, k, \lambda, \mu)$ is a k-regular graph on n vertices for which any pair of adjacent (respectively non-adjacent) vertices have exactly λ (resp. μ) neighbours in common.

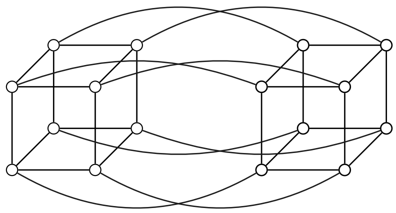


Figure 4.11: The hypercube $\mathcal{H}_4$ of dimension 4.

For instance, the cycle $\mathcal{C}_4$ is a strongly regular graph with parameters $n = p$, $k = 2$, $\lambda = 0$ and $\mu = 2$, but the cycle $\mathcal{C}_6$ is not a strongly regular graph as the number of common neighbours of two non-adjacent vertices is not constant.

The four parameters of a strongly regular graph are related in the following way.

Proposition 4.14.

Let G be a $\mathrm{srg}(n, k, \lambda, \mu)$. We have

$$(n - k - 1)\mu = k(k - \lambda - 1). \tag{4.1}$$

Proof. Let $G = (V, E)$ be a $\mathrm{srg}(n, k, \lambda, \mu)$ and let u be a vertex of V. We consider the partition of $V \setminus \{u\}$ between the neighbours $N(u)$ and the non-neighbours $V \setminus N[u]$ of u. By definition, $\#N(u) = k$ and $\#(V \setminus N[u]) = n - 1 - k$. We now count the number of edges between $N(u)$ and $V \setminus N[u]$ (Figure 4.12). For any $v \in N(u)$, there are λ edges to vertices of $N(u)$ as u and v are adjacent. Hence, there are $k - 1 - \lambda$ edges between v and vertices of $V \setminus N[u]$. Similarly, for any $w \in V \setminus N[u]$, there are μ edges to vertices of $N(u)$ as u and w are not adjacent. Therefore, counting the number of edges between $N(u)$ and $V \setminus N[u]$ in two different ways, we obtain the required equality. □

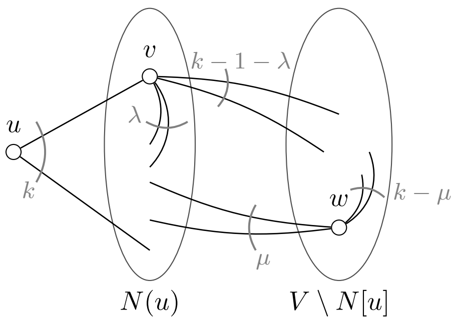


Figure 4.12: A strongly regular graph $\mathrm{srg}(n, k, \lambda, \mu)$ where the number of edges incident with a given vertex is indicated in gray.

The complement $\overline{G}$ of a strongly regular graph G is still a strongly regular graph and has parameters $\text{srg}(n, n-1-k, n-2-2k+\mu, n-2k+\lambda)$. Indeed, by definition $V(\overline{G}) = V(G)$ and the degree of any vertex v in the complement is clearly $(n-1)-k$. Consider two adjacent vertices u and v in G (Figure 4.13), they have λ common neighbours. Let S denote the set $V(G) \setminus (N[u] \cup N[v])$. In $\overline{G}$, u and v are not adjacent and their common neighbours are the vertices of S which has cardinality equal to $n-(k+1)-(k-1-\lambda) = n-2k+\lambda$. Similarly, if u and v are non-adjacent vertices in G, then they have μ common neighbours. We set $S = V(G) \setminus (N[u] \cup N[v])$ again. In $\overline{G}$, u and v are adjacent and the number of their common neighbours is $|S| = n-(k+1)-(k+1-\mu) = n-2-2k-\mu$.

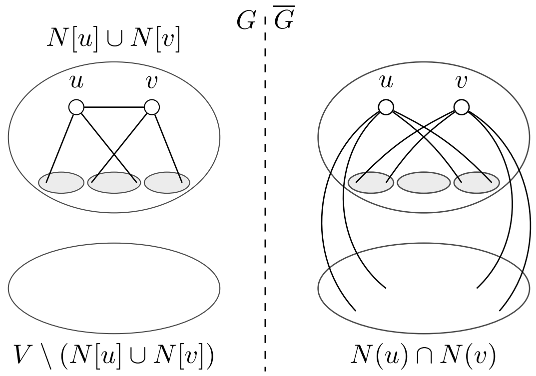


Figure 4.13: From a strongly regular graph $\text{srg}(n, k, \lambda, \mu)$, denoted by G, to its complement $\overline{G}$. Focus on two adjacent vertices.

Definition 4.15. A strongly regular graph is *primitive* if the graph and its complement are connected.

Example 4.16. Let G be a $\text{srg}(n, k, \lambda, \mu)$. A trivial non primitive case is given by $\mu = 0$. Indeed, if $\mu = 0$, then it is the disjoint union of complete graphs on $k+1$ vertices. In particular, $\lambda = k-1$. Indeed, let u and v be two adjacent vertices of G. Assume there exists w adjacent to u but not to v. Then v and w are non-adjacent vertices with a common neighbours, a contradiction.

Similarly, if $\mu = k$, then G is a complete multipartite graph. Indeed, the relation of being non-adjacent to u is an equivalence relation since two non-adjacent vertices have exactly the same open neighbourhood. Hence G is a complete multipartite graph. Necessarily, all the parts have the same size, $n-k$. Note that the complement of G corresponds to the first example.

The two previous examples are the only non primitive graphs.

Lemma 4.17.

Let G be a strongly regular graph. G is primitive if and only if $\mu \notin \{0, k\}$. In particular, all primitive strongly regular graphs have diameter 2.

Proof. As shown in Example 4.16, if $\mu \in \{0, k\}$ then G is not primitive. If $\mu \neq 0$, then two non-adjacent vertices have at least one vertex in common. Hence the diameter of G is two and in particular, G is connected. Assume now that $\mu \neq k$. By Equation (4.1), the value of μ for the complement of G, $n - 2k + \lambda$, is not 0. As before, it means that the complement of G has diameter 2 and is connected. □

Proposition 4.18.

Let G be a primitive strongly regular graph $\mathrm{srg}(n, k, \lambda, \mu)$ on n vertices, then $k \geq \sqrt{n-1}$ and the smallest symmetric difference satisfies $d > \sqrt{n} - 3$.

Proof. Since G is primitive, by Lemma 4.17, it has diameter 2. Thus there are at most $1 + k + k(k-1)$ vertices in G. Hence $n \leq 1 + k^2$ and we get the upper bound on k.

To prove the second inequality, we use a result of Babai [Bab80]: for every pair of vertices u, v of a primitive strongly regular graph, one has $|N(u)\Delta N(v)| > \sqrt{n} - 1$. If u and v are adjacent, then $|N[u]\Delta N[v]| = |N(u)\Delta N(v)| - 2$, whereas if u and v are non-adjacent, $|N[u]\Delta N[v]| = |N(u)\Delta N(v)| + 2$. Hence $d > \sqrt{n} - 3$.

□

4.3.4 Infinite graphs

The *infinite path*, denoted by P_∞, has vertex-set $\mathbb{Z}$ and two vertices u and v are adjacent if $|u - v| = 1$.

The *infinite square grid*, or simply *infinite grid*, is the Cartesian product of two infinite paths. We can also view the infinite grid as $\mathbb{Z}^2$. The vertices are all pairs of integers and two vertices (x_1, x_2) and (y_1, y_2) are adjacent if $|x_1 - y_1| + |x_2 - y_2| = 1$. The infinite grid is a 4-regular graph, i.e., every vertex has 4 neighbours. Let the sets

$$L_e = \{(x_1, x_2) \in \mathbb{Z}^2 \mid x_1 + x_2 = 0 \pmod 2\}$$

and

$$L_o = \{(x_1, x_2) \in \mathbb{Z}^2 \mid x_1 + x_2 = 1 \pmod 2\}$$

denote the *even* and *odd sub-lattices* of $\mathbb{Z}^2$. We have $\mathbb{Z}^2 = L_e \cup L_o$. Sets of the type $\{(x_1, x_1 + c) \mid x_1 \in \mathbb{Z}\}$ and $\{(x_1, -x_1 + c) \mid x_1 \in \mathbb{Z}\}$ with $c \in \mathbb{Z}$ are called *diagonals* of $\mathbb{Z}^2$.

The *infinite king lattice* has also all pairs of integers as vertices, but two vertices (x_1, x_2) and (y_1, y_2) are adjacent if $|x_1 - y_1| \leq 1$ and $|x_2 - y_2| \leq 1$. Otherwise stated, the neighbourhood of a vertex corresponds to the possible moves of a king on a chessboard. Figure 4.14 displays portions of the infinite square grid and the infinite king lattice. Observe that in Figure 4.14(b), the white vertices correspond to vertices of the odd sub-lattice.

4.3.5 Hypergraphs

A *hypergraph* is a couple $(V, \mathcal{E})$ where V is a set and $\mathcal{E}$ is a subset of the power set of V. The elements of V are called *vertices* and the elements of $\mathcal{E}$ are called *hyperedges*. A simple undirected graph can be viewed as a hypergraph with all its hyperedges of cardinality 2. An example of hypergraph $(V, \mathcal{E})$ is given in Figure 4.15 where the set of vertices is $V = \{v_0, \ldots, v_7\}$ and the set of hyperedges is

$$\mathcal{E} = \{\{v_{0,3}, v_5, v_6\}, \{v_0, v_5\}, \{v_2, v_4, v_6\}, \{v_2, v_4, v_7\}\}.$$

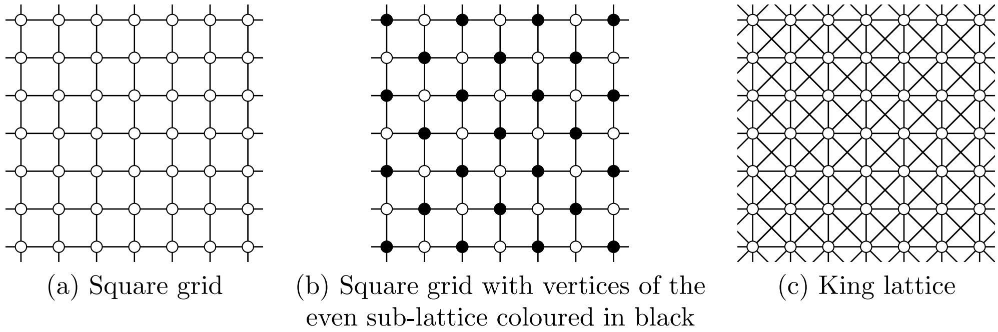


Figure 4.14: Portions of the infinite square grid and of the infinite king lattice.

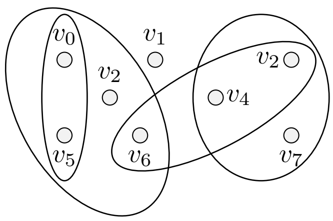


Figure 4.15: Example of hypergraph.

4.4 Colouring of graphs

A *colouring* of the vertices of a graph G is a map $c : V(G) \to \mathbb{N}$. For a vertex u, $c(u)$ is the *colour* of u. A k-colouring is a colouring using only k colours. A colouring of the vertices of a graph is *monochromatic* if all vertices have the same color. In this thesis, we are interested in 2-colourings, $c : V(G) \to \{0, 1\}$. Hence we identify the colour 0 with white and 1 with black. Observe that a 2-colouring c of $V(G)$ can be translated in terms of considering a subset C of $V(G)$:

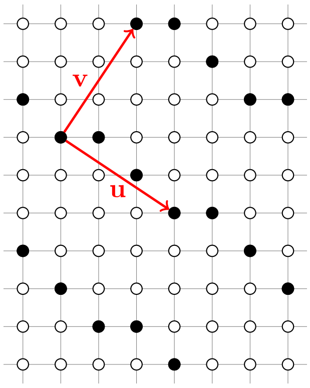
$$c(v) = 1 \text{ if and only if } v \in C.$$

A 2-colouring of P_∞ is *periodic* if there exists a positive integer p such that $c(x) = c(x+p)$ for all $x \in \mathbb{Z}$. In that case, the smallest p satisfying the previous condition is called the *period* and any pattern (i.e., sequence of colours) of length p appearing in the colouring is a *pattern period*. The notion of periodicity in infinite paths is similar to the notion of periodicity in bi-infinite words, i.e., sequences indexed by $\mathbb{Z}$. A 2-colouring c of P_∞ is *anti-periodic* of anti-period p is $c(x) = 1 - c(x+p)$ for all $x \in \mathbb{Z}$. Hence such colourings are periodic of period $2p$. As each line or row of the infinite grid is isomorphic to P_∞, we can define these notions analogously for lines and rows of $\mathbb{Z}^2$. Similarly, we can adapt these definitions to the case of cycles by working modulo p. Note that a 1-periodic colouring is simply a monochromatic colouring. A 1-anti-periodic colouring is called *alternate*.

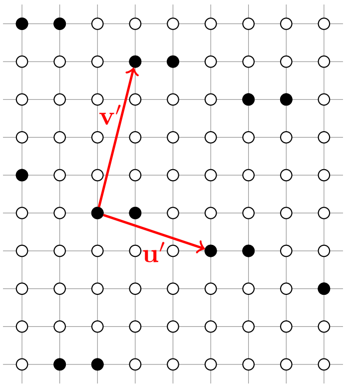
Consider now a colouring c of the vertices of the infinite grid $\mathbb{Z}^2$. We say that c is *periodic* if there exist two non-proportional vectors $\mathbf{u}, \mathbf{v}$ such that $c(\mathbf{x}+\mathbf{u}) = c(\mathbf{x}) = c(\mathbf{x}+\mathbf{v})$ for all $\mathbf{x} \in \mathbb{Z}^2$. In particular, if c is a periodic colouring of $\mathbb{Z}^2$, then there exist two integers m and n distinct from 0 such that $c(\mathbf{x}+(m,0)) = c(\mathbf{x}) = c(\mathbf{x}+(0,n))$ for all $\mathbf{x} \in \mathbb{Z}^2$. Otherwise stated, if c is periodic, then each line and each row are coloured periodically.

Example 4.19. The portions of two infinite grids represented in Figure 4.16 are coloured with a periodic 2-colouring. For the colouring c given in Figure 4.16(a), the vectors $\mathbf{u} = (3,-2)$ and $\mathbf{v} = (2,3)$ are non-proportional and such that $c(\mathbf{x}+\mathbf{u}) = c(\mathbf{x}) = c(\mathbf{x}+\mathbf{v})$ for any $\mathbf{x} \in \mathbb{Z}^2$. To find the periods of each line and each row, observe that $3\mathbf{u}+2\mathbf{v} = (13,0)$ and $-2\mathbf{u}+3\mathbf{v} = (0,13)$. Hence, $c(\mathbf{x}) = c(\mathbf{x}+(13,0)) = c(\mathbf{x}+(0,13))$ and the periods of each line and of each row are equal to 13.

Now consider the colouring c' given in Figure 4.16(b). We have $\mathbf{u}' = (3,-1)$ and $\mathbf{v}' = (1,4)$. We obtain again $c'(\mathbf{x}) = c'(\mathbf{x}+(13,0)) = c'(\mathbf{x}+(0,13))$ for any $\mathbf{x} \in \mathbb{Z}^2$.



(a) $\mathbf{u} = (3,-2)$, $\mathbf{v} = (2,3)$



(b) $\mathbf{u}' = (3,-1)$, $\mathbf{v}' = (1,4)$

Figure 4.16: Portions of two infinite grids, both coloured periodically.

If c is a colouring of $\mathbb{Z}^2$ such that the even and odd sublattices are the disjoint union of monochromatic diagonals, then c is called a *diagonal colouring*. Observe that the monochromatic diagonals of the even sublattice do not have to be parallel to the monochromatic diagonals of the odd sublattice (Figure 4.17). A diagonal colouring c of $\mathbb{Z}^2$ is *p-periodic* (respectively *p-anti-periodic*) if horizontal lines are coloured p-periodically (resp. p-anti-periodically).

For any 2-colouring c, we write $\overline{c}$ for the *complement* of c, that is, the colouring obtained from c by changing black into white and vice versa. Hence, we have $\overline{c}(u) = 1 - c(u)$ for any vertex u. Observe that this notion is similar to the complement of binary words defined in Section 1.1.

4.5 Covering problems

Covering and packing problems are traditional issues in mathematics [CHLL97]. A natural packing problem in the n-dimensional euclidean space is to ask for the maximal number of identical non-intersecting spheres in a large volume. Conversely, a covering problem in the

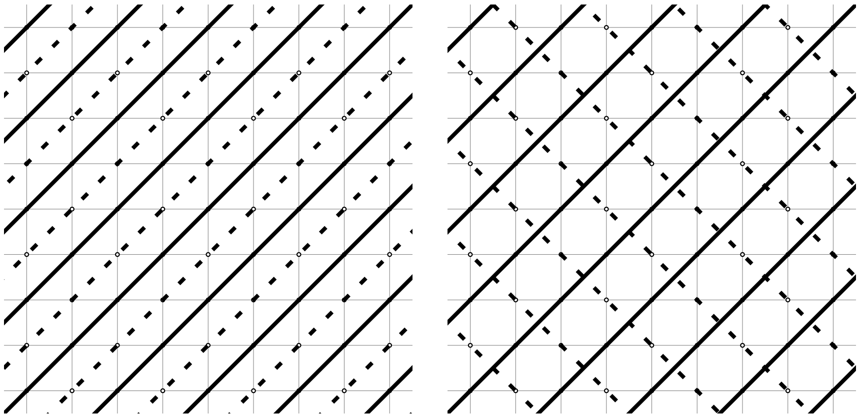


Figure 4.17: Schemes of diagonal colourings of the infinite grid with respectively parallel and non-parallel monochromatic diagonals.

euclidean space asks for the minimal number of identical spheres needed to cover a large volume. See [CHLL97] for many bibliographic pointers.

The same issues can be considered in graphs. Given an r and a graph, a packing problem is for instance to determine the maximal number of non-intersecting r-balls that can be placed in the graph; a covering problem is for example to determine the minimal number of r-balls that can be placed in such a way that every vertex of the graph is contained in at least one of them. Packing problems are fundamental in "error correction" while covering problems have application in mobile network.

We consider in this section covering problems that satisfy special conditions. Firstly, we are interested in using a covering with balls of radius 1 that permit us to identify each vertex of a given graph. Secondly, we focus on covering problems that satisfy multiplicity conditions. In each subsection, we present an application of these coverings.

4.5.1 Identifying codes

Given a discrete structure on a set of elements, a natural question is to be able to locate efficiently the elements using the structure. If the elements are the vertices of a graph, one can use the neighbourhoods of the elements to locate them. In this context, Karpovsky et al. [KCL98] have introduced the notion of identifying codes in 1998.

Definition 4.20. A subset of vertices S is a *dominating set* if each vertex is either in S or adjacent to a vertex in S. In other words, for every vertex u, $S \cap N[u]$ is non-empty. A vertex c *separates* two vertices u and v if exactly one vertex among u and v is in the closed neighbourhood of c. In other words, $c \in N[u]\Delta N[v]$ where Δ denotes the symmetric difference of sets. A subset of vertices S is a *separating set* if it separates every pair of vertices of the graph. A subset of vertices C is an *identifying code* if it is both a dominating and separating set. In other words, the set $N[u] \cap C$ is non-empty and uniquely determines u.

Example 4.21. The set $C = \{v_0, v_2, v_3, v_5\}$ is an identifying code of the graph given in Figure 4.18. Indeed, we can check that each subset $N[u] \cap C$ appearing in Table 4.1 is non-empty and unique.

u	v_0	v_1	v_2	v_3	v_4	v_5	v_6
$N[u] \cap C$	$\{v_0\}$	$\{v_0, v_2\}$	$\{v_2, v_3, v_5\}$	$\{v_2, v_3\}$	$\{v_0, v_5\}$	$\{v_2, v_5\}$	$\{v_5\}$

Table 4.1: Set of neighbours within a set C.

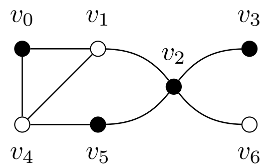


Figure 4.18: The set of black vertices is an identifying code of the graph.

Initially, identifying codes have been introduced to model fault-diagnosis in multiprocessor systems [KCL98]. A multiprocessor system is represented by a graph where the vertices are the processors. Assume that a processor can test if a neighbour processor is faulty and only returns a binary value. For instance, a processor returns 0 if no faults were detected and 1 otherwise. The problem is to find a subset C of processors such that

- if all processors of C return the value 0, it means that no processors are faulty,
- if at least one processor of C returns the value 1, then there is a faulty processor and we can uniquely determine which one it is.

If we suppose that at any moment, there is at most one faulty processor, then the wanted set C corresponds exactly to an identifying code.

Indeed, the first condition ensures that if there is a faulty processor, then it will be detected. In terms of graphs, it means that C is a dominating set. The second condition, about the unique localisation of the processor, is equivalent to the condition that C is a separating set of the graph. Later other applications were discovered such as the design of emergency sensor networks in facilities [UTS04].

Remark 4.22. The identifying code problem in graphs is equivalent to a covering problem in hypergraphs. Let $G = (V, E)$ be a graph and let $\mathcal{H} = (V, \mathcal{E})$ be the hypergraph with hyperedge set

$$\mathcal{E} = \{N[u] \mid u \in V\} \cup \{N[u] \Delta N[v] \mid u \neq v \in V\}.$$

A subset C of vertices is an identifying code of G if and only if C intersects all the hyperedges.

For instance, consider the identifying code $\{v_0, v_2, v_3, v_5\}$ of the graph depicted in Figure 4.18. We can check that this set intersects all the hyperedges of the corresponding hypergraph (Table 4.2).

There exists an identifying code in G if and only if G does not have two vertices u and v with $N[u] = N[v]$. We say that two such vertices u and v are *twin vertices*. For instance, a cycle $\mathcal{C}_4$ of length 4 has two pairs of twin vertices (Figure 4.19). In the sequel we only consider twin-free graphs. The size of a minimal identifying code of G is denoted by $\gamma^{\mathrm{ID}}(G)$. We have the following general bounds.

$N[v_0]$	v_0	v_1			v_4			$N[v_1]\Delta N[v_3]$	v_0	v_1		v_3	v_4	v_5	
$N[v_1]$	v_0	v_1			v_4	v_5		$N[v_1]\Delta N[v_4]$			v_2				
$N[v_2]$		v_1	v_2	v_3		v_5	v_6	$N[v_1]\Delta N[v_5]$	v_0						
$N[v_3]$			v_2	v_3				$N[v_1]\Delta N[v_6]$	v_0	v_1			v_4	v_5	v_6
$N[v_4]$	v_0	v_1			v_4	v_5		$N[v_2]\Delta N[v_3]$		v_1				v_5	v_6
$N[v_5]$		v_1	v_2		v_4	v_5		$N[v_2]\Delta N[v_4]$	v_0		v_2	v_3	v_4		v_6
$N[v_6]$			v_2				v_6	$N[v_2]\Delta N[v_5]$	v_0			v_3	v_4		v_6
$N[v_0]\Delta N[v_1]$			v_2			v_5		$N[v_2]\Delta N[v_6]$		v_1		v_3		v_5	
$N[v_0]\Delta N[v_2]$	v_0				v_4			$N[v_3]\Delta N[v_4]$	v_0	v_1	v_2	v_3	v_4	v_5	
$N[v_0]\Delta N[v_3]$	v_0	v_1	v_2	v_3	v_4			$N[v_3]\Delta N[v_5]$		v_1		v_3	v_4	v_5	
$N[v_0]\Delta N[v_4]$						v_5		$N[v_3]\Delta N[v_6]$				v_3			v_6
$N[v_0]\Delta N[v_5]$	v_0		v_2			v_5		$N[v_4]\Delta N[v_5]$			v_2				
$N[v_0]\Delta N[v_6]$	v_0	v_1	v_2		v_4		v_6	$N[v_4]\Delta N[v_6]$	v_0	v_1	v_2		v_4	v_5	v_6
$N[v_1]\Delta N[v_2]$	v_0				v_4			$N[v_5]\Delta N[v_6]$		v_1			v_4	v_5	v_6

Table 4.2: Each row corresponds to an hyperedge of $\mathcal{H}$ and the gray columns correspond to vertices of the set $\{v_0, v_2, v_3, v_5\}$.

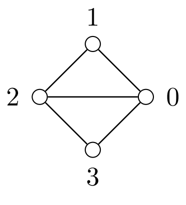


Figure 4.19: The vertices 0 and 2 are twin vertices.

Proposition 4.23. [GM07, KCL98]

Let G be a twin-free graph with at least one edge. We have

$$\lceil \log_2(|V|+1) \rceil \leq \gamma^{\mathrm{ID}}(G) \leq |V|-1.$$

The lower bound can be found by considering that in an identifying code C of size $\gamma^{\mathrm{ID}}(G)$, the sets $N[u] \cap C$ are all distinct and non-empty subsets of a set of size $\gamma^{\mathrm{ID}}(G)$. Both bounds are tight and graphs reaching the lower bound are described in [Mon06] whereas graphs reaching the upper bound are characterized in [FGK+11]. For instance, Figure 4.20 represents one graph reaching the lower bound and another graph reaching the upper bound.

$\gamma^{\mathrm{ID}}(G) = 3 = \lceil \log_2(|V|+1) \rceil$ $\qquad$ $\gamma^{\mathrm{ID}}(G) = 6 = |V|-1$

Figure 4.20: Two graphs reaching the lower and upper bounds on $\gamma^{\mathrm{ID}}(G)$ in terms of the number of vertices.

When the maximum degree of the graph is small enough, the following lower bound is more precise than the previous one.

Proposition 4.24. [KCL98]

Let G be a graph of maximum degree k. We have

$$\gamma^{\mathrm{ID}}(G) \geq \frac{2|V|}{k+1}.$$

Karpovsky et al. [KCL98] proved this bound using a discharging method that is illustrated in Figure 4.21. For a fixed subset C of vertices of a graph, each vertex receives a charge 1 at the beginning. Then each vertex v gives to the vertices in $N[v] \cap C$ the charge $\frac{1}{|N[v] \cap C|}$. After this process, only vertices of C have a positive charge and the total charge is still $|V|$. We use the same method to obtain a tighter bound which we need when $\gamma^{\mathrm{ID}}(G)$ is smaller than the maximum degree of the graph.

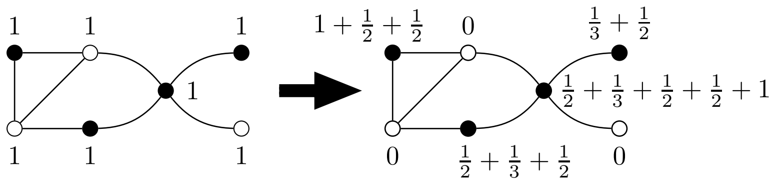


Figure 4.21: Charges of the graph before and after applying the discharging method.

Proposition 4.25.

Let $G = (V, E)$ be a twin-free graph of maximum degree k and C an identifying code of G with $k \geq |C| + 1$. We have

$$|V| \leq \frac{|C|^2}{6} + \frac{(2k+5)|C|}{6}.$$

Proof. Let $G = (V, E)$ be a twin-free graph of maximum degree k. Assume that C is an identifying code of G with $k \geq |C| + 1$. We use the same discharging method as Karpovsky et al. in [KCL98]. Each vertex receives a charge 1 at the beginning. Then each vertex v gives to the vertices in $N[v] \cap C$ the charge $\frac{1}{|N[v] \cap C|}$. After this process, only vertices of C have a positive charge and the total charge is still $|V|$.

Let $c \in C$. Let V_i be the set of vertices of $N[c]$ with exactly i neighbours in C. Necessarily $|V_1| \leq 1$ since vertices in V_1 have only c in their neighbourhood. We have $|V_2| \leq |C| - 1$. Indeed, a vertex of V_2 has c in its neighbourhood and a unique additional vertex of the code. But all the additional code neighbours of elements of V_2 must be different, hence there are at most $|C| - 1$ vertices in V_2. Finally, there are $k + 1 - |V_1| - |V_2|$ other vertices giving charge at most $1/3$. Therefore, c receives a charge at most equal to

$$|V_1| + \frac{|V_2|}{2} + \frac{k + 1 - |V_1| - |V_2|}{3} \leq 1 + \frac{|C| - 1}{2} + \frac{k - |C| + 1}{3} = \frac{|C|}{6} + \frac{2k+5}{6}.$$

Hence the total charge $|V|$ is at most $\frac{|C|^2}{6} + \frac{(2k+5)|C|}{6}$. □

Remark 4.26. The problem of finding a minimal identifying code in a graph G can be expressed as a hitting set problem. Indeed an identifying code is a subset of V that intersects all the sets $N[u]$ and $N[u] \Delta N[v]$ for $u, v \in V$. In other words, the problem of finding a minimal identifying code is equivalent to the following linear integer program P_G.

$$\begin{array}{llllll} \text{Minimize} & \displaystyle\sum_{x_u \in V} x_u & & & & \\ \text{such that} & \displaystyle\sum_{w \in N[u]} x_w & \geq 1 & & \forall u \in V & \text{(domination)} \\ & \displaystyle\sum_{w \in N[u] \Delta N[v]} x_w & \geq 1 & \forall u, v \in V, u \neq v & & \text{(separation)} \\ & x_u \in \{0,1\} & & & \forall u \in V & \end{array}$$

The concept of identifying codes is related to other concepts such as locating-dominating sets [Sla87, Sla88] and resolving sets [Bab80, Sla75]. A *locating-dominating set* is a dominating set S that separates the pairs of vertices that are not in S [Sla87, Sla88]. The size of a minimal locating-dominating set of G is denoted by $\gamma^{\mathrm{LD}}(G)$. Note that every graph admits a locating-dominating set since the whole set of vertices is always a locating-dominating set. An identifying code is always a locating-dominating set and one can get an identifying code from a locating-dominating set by adding at most $\gamma^{\mathrm{LD}}(G)$ vertices. Therefore we have the following relations between $\gamma^{\mathrm{LD}}(G)$ and $\gamma^{\mathrm{ID}}(G)$.

Proposition 4.27. [GKM08]

Let G be a twin-free graph. We have

$$\gamma^{\mathrm{LD}}(G) \leq \gamma^{\mathrm{ID}}(G) \leq 2\gamma^{\mathrm{LD}}(G).$$

Example 4.28. Consider the graph with vertex set $V = \{v_0, \ldots, v_6\}$ and edge set

$$E = \{v_0v_1, v_0v_4, v_1v_2, v_1v_4, v_2v_3, v_2v_5, v_2v_6, v_4v_5\}.$$

The set $C = \{v_0, v_3, v_4, v_6\}$ is a locating-dominating set of the graph (Figure 4.22(a)) but it is not an identifying code since $N[v_0] \cap C = \{v_0, v_4\} = N[v_4] \cap C$. To obtain an identifying code from C, it suffices to add the vertices v_2 and v_5 (Figure 4.22(b)). It is not the only solution to obtain an identifying code. For instance, we can add the vertices v_1, v_2 instead of v_2, v_5.

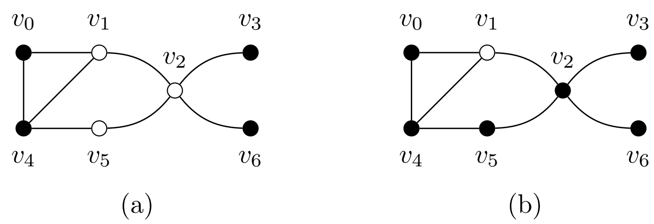
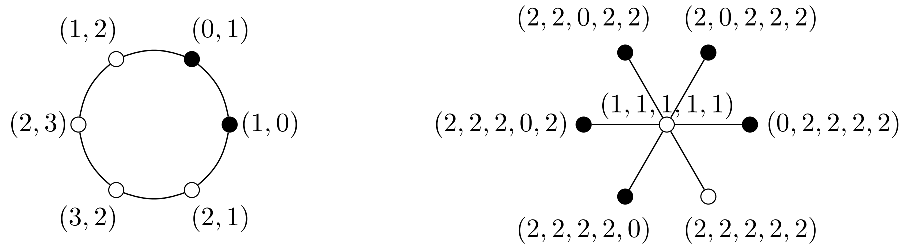


Figure 4.22: In (a), the set of black vertices is a locating-dominating set but not an identifying code. In (b), the set of black vertices is a locating-dominating set and an identifying code.

A *resolving set* is a subset of vertices S such that for every pair of vertices u and v, there exists a vertex x in S that satisfies $d(x,u) \neq d(x,v)$ [Bab80, Sla75]. The smallest size of a resolving set of G is called the *metric dimension* and is denoted by $\beta(G)$. Two examples of resolving sets are given in Figure 4.23.



$\beta(G) = 2$ $\qquad\qquad$ $\beta(G) = 5$

Figure 4.23: For each graph, the black vertices form a resolving set. For any vertex, the distance between each black vertex and itself is indicated.

A locating-dominating set is always a resolving set and so $\beta(G) \leq \gamma^{\text{LD}}(G)$. When the diameter of the graph is 2, the reverse is almost true: adding a vertex to a resolving set gives a locating dominating set. For instance, in both graphs depicted in Figure 4.23, one has $\gamma^{\text{LD}}(G) = \beta(G)$.

Proposition 4.29.

Let G be a graph of diameter 2. We have

$$\beta(G) \leq \gamma^{\text{LD}}(G) \leq \beta(G) + 1.$$

Proof. The first inequality is true for any graph since a locating-dominating set is a resolving set. Let now S be a resolving set of a graph G of diameter 2. We order the vertices of $S = \{x_1, ..., x_s\}$. For every vertex u, let $L(u) = (d(u,x_1), ..., d(u,x_s))$ be the distance vector to vertices of S. Since S is a resolving set, all the vectors $L(u)$ are distinct. Since the diameter is 2, $L(u) \in \{0,1,2\}^s$. But at most one vertex u_0 can have $L(u_0) = (2,2,...,2)$, hence all vertices except u_0 are dominated by a vertex of S. Therefore, the set $S' = S \cup \{u_0\}$ is a dominating set. Let u be a vertex not in S'. It has only values 1 and 2 in its vector $L(u)$ and the set $N[u] \cap S$ is given by the value 1 in $L(u)$. Hence all the sets $N[u] \cap S$ for $u \notin S'$ are distinct. Therefore, all the sets $N[u] \cap S'$ are also distinct for $u \notin S'$ and S' is a locating-dominating set. In particular $\gamma^{\text{LD}}(G) \leq \beta(G) + 1$. $\square$

Proposition 4.27 together with Proposition 4.29 gives a relation between $\gamma^{\text{ID}}(G)$ and the metric dimension in graphs of diameter 2. In particular, they have the same order and let us derive results for identifying codes from results for resolving sets.

Corollary 4.30.

Let G be a twin-free graph of diameter 2. We have

$$\beta(G) \leq \gamma^{\mathrm{ID}}(G) \leq 2\beta(G) + 2.$$

4.5.2 (r, a, b)-covering codes

One of the motivations for studying covering codes in special graphs is a network communication problem. Consider for example a system of transmitting stations for cellular phone network. In Figure 4.24, we locate transmitting stations in the vertices of the corresponding graphs such that any two stations are at distance at least $r+1$ from each other (to avoid interference) but any other vertex is within reaching distance r from 2 transmitting stations (to guarantee a good quality of transmission). Such coverings are called (r, a, b)-covering codes where a is the number of transmitting stations within a distance r from a given transmitting station ($a = 1$ in the example above) and b is the number of transmitting stations within a distance r from a given vertex that is not a transmitting station ($b = 2$ in the example above).

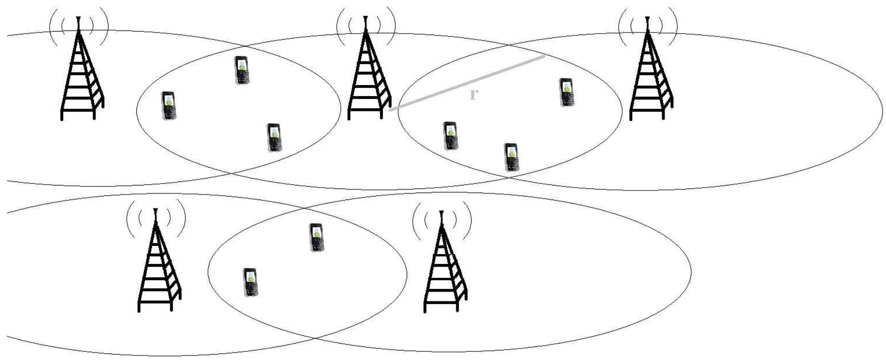


Figure 4.24: This placement of the transmitting stations with respect to the cellular phones is an example of an $(r, 1, 2)$-covering code.

Definition 4.31. Let $G = (V, E)$ be a graph and r, a, b be positive integers. A set $S \subseteq V$ of vertices is an (r, a, b)-*covering code* or simply (r, a, b)-*code* if every element of S belongs to exactly a balls of radius r with elements of S as centers and every element of $V \setminus S$ belongs to exactly b balls of radius r with elements of S as centers. In other words, (r, a, b)-covering code is such that for a vertex u,

$$\#\{B_r(v) \mid u \in B_r(v),\ v \in S\} = \begin{cases} a & \text{if } u \in S \\ b & \text{if } u \notin S. \end{cases}$$

Such codes are also known as (r, a, b)-isotropic colourings [Axe03] or as perfect colourings [Puz08].

We can view an (r, a, b)-code as a particular colouring c with two colors, black and white, where the black vertices are the elements of the code. Hence, the colouring c is such that an r-ball with a black (respectively white) vertex as center contain exactly a (resp. b) black vertices.

Example 4.32. The periodic colouring given in Figure 4.16(a) is a $(2, 3, 2)$-code of the infinite grid. Since the colouring is periodic, it suffices to check for a finite number of vertices that each balls of radius contain either 3 black vertices if the center is a black vertex, or 2 black vertices if the center is a white vertex. In this particular example, 15 well-chosen vertices are enough. Similarly, the periodic colouring of $\mathbb{Z}^2$ given in Figure 4.16(b) is a $(3, 3, 4)$-code.

Surprisingly, (r, a, b)-covering codes do not exist for every value of the parameters even in the particular case of the infinite square grid. For instance, $(r, 1, 3)$-covering code does not exist [Axe03].

The notion of (r, a, b)-codes generalizes the notion of domination and perfect codes in graphs. Perfect codes were introduced in terms of graphs by Biggs [Big73]. An *r-perfect code* of a graph $G = (V, E)$ is a subset $C \subseteq V$ with the property that each vertex is within distance r of exactly one vertex of C. In other words, the balls of radius r with elements of C as centers form a partition of V. Hence, an r-perfect code is an $(r, 1, 1)$-code. If $r = 1$, then a 1-perfect code is a dominating set with no adjacent vertices. Figure 4.25 depicts two examples of r-perfect codes of the Cartesian product of two paths P_4.

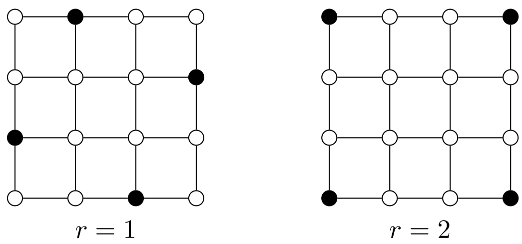


Figure 4.25: The set of black vertices is an r-perfect code, i.e., $(r, 1, 1)$-code, of $P_4 \square P_4$.

Kratochvíl [Kra88] showed that the problem of finding an r-perfect code in graphs is NP-complete. Moreover, this problem is still NP-complete in the particular case of bipartite graphs with maximum degree three. For more information about perfect codes, see [CHLL97, Chapter 11].

Perfect codes have also been studied in infinite graphs. For example, Golomb and Welsh [GW68, GW70] considered the multidimensional rectangular grid $\mathbb{Z}^d$. They proved the existence of 1-perfect codes, i.e., $(1, 1, 1)$-codes, in $\mathbb{Z}^d$. Such codes can be considered as periodic tilings of the grid $\mathbb{Z}^n$ by balls of radius 1. Moreover, the authors conjectured that there do not exist r-perfect codes with $r > 1$ in $\mathbb{Z}^d$ [GW68, GW70].

The (r, a, b)-codes have already been studied in some graphs under the names of *weighted covering codes* by Cohen et al. [CHLM95]. Their work corresponds to a study of these codes in the Hamming metric. For a subset C of vertices, they attach weights to different layers of the Hamming sphere and they consider weighted spheres centred at vertices of C. If several such spheres intersect in a vertex, they define the density of each vertex as the sum of the weights of the corresponding layers. The set C is called a *weighted covering* if the density at each vertex is at least one. When the density is exactly equal to one for all vertices, then C is called a *perfect weighted covering*. If the radius is equal to 1, a $(1, a, b)$-code is exactly a perfect weighted covering of radius one with weight $\left(\frac{b-a+1}{b}, \frac{1}{b}\right)$. For more details see [CHLL97, Chapter 13].

While Cohen et al. [CHLM95] studied weighted codes in Hamming metric, Telle considered a particular case of these codes in graphs in general [Tel94]. For a subset C of vertices, he defines the state of a vertex $u \in C$ by

$$state(u) = \begin{cases} \sigma_i & \text{if } u \in C \text{ and } |N(u) \cap C| = i \\ \rho_i & \text{if } u \notin C \text{ and } |N(u) \cap C| = i. \end{cases}$$

Then many properties of vertex subsets can be defined by allowing only a specific set L of states. For instance, the set C is a dominating set if the state ρ_0 is not allowed. In this setting, $(1, a, b)$-codes are equivalent to $[\sigma_{a-1}, \rho_b]$-dominating sets. Telle proved [Tel94] that the following decidability problem was NP-complete: "Is it possible to decide whether a graph has an $[\sigma_a, \rho_b]$-dominating set ?". The problem is still NP-complete when restricted to planar bipartite graphs of maximum degree three.

The particular case where the radius is 1 has been studied a lot. In the multidimensional grid, which corresponds to the Lee metric with an infinite alphabet, $(1, a, b)$-codes were studied by Dorbec et al. [DGHM09] and Gravier et al. [GMP99]. For instance, the existence of $(1, 2, 1)$-codes[1] in $\mathbb{Z}^d$ is proved in both papers [DGHM09, GMP99]. In [DGHM09, Theorem 4], Dorbec et al. present a method to construct $(1, a, b)$-codes in $\mathbb{Z}^d$. This method is based on a one-dimensional pattern of finite length that is extended by translations to colour $\mathbb{Z}^d$. Hence, the code obtained satisfies periodic properties.

Theorem 4.33. [DGHM09]

Assume that $1 \leq k \leq n, 1 \leq d$ and

$$A = \{a_1, a_2, \ldots, a_k\} \subseteq \mathbb{Z}_n \text{ (where } a_i \neq a_j, \text{ when } i \neq j$$

and $w_1, \ldots, w_d$ are (not necessarily distinct) elements of $\mathbb{Z}_n$. Consider the sums $a_i + w$ and the differences $a_j - w_j$. If these $2kd$ elements take each value in A exactly a times and each value in $\mathbb{Z}_n \setminus A$ exactly b times, then the set

$$C = \{(x_1, \ldots, x_d) \in \mathbb{Z}^d \mid x_1 w_1 + \cdots + x_d w_d \in A\}$$

is an $(1, a+1, b)$-code of $\mathbb{Z}^d$.

[1] The reader may notice a distinction of notation between this thesis and [DGHM09]. They write (a, b)-codes for what we denote $(1, a+1, b)$-code as they consider open neighbourhoods and we consider closed neighbourhoods.

Example 4.34 ([Dor07])**.** We present here an illustration of the method described in the previous theorem [DGHM09]. For instance, consider $n = 15$ and the set $A = \{1, 2, 3\}$ that corresponds to the pattern

$$\bullet\bullet\bullet\circ\circ\circ\circ\circ\circ\circ\circ\circ\circ\circ\circ,$$

and $w_1 = 0, w_2 = 2, w_3 = 4, w_4 = 5, w_5 = 7$. Table 4.3 shows the existence of a $(1, 5, 2)$-code C in $\mathbb{Z}^6$.

A	●	●	●	○	○	○	○	○	○	○	○	○	○	○	○
$\|N[u] \cap A\|$	2	3	2	1	0	0	0	0	0	0	0	0	0	0	1
$w_1 = 0$	●	●	●	○	○	○	○	○	○	○	○	○	○	○	○
$-w_1 = 0$	●	●	●	○	○	○	○	○	○	○	○	○	○	○	○
$w_2 = 2$	○	○	●	●	●	○	○	○	○	○	○	○	○	○	○
$-w_2 = -2$	●	○	○	○	○	○	○	○	○	○	○	○	○	●	●
$w_3 = 4$	○	○	○	○	●	●	●	○	○	○	○	○	○	○	○
$-w_3 = -4$	○	○	○	○	○	○	○	○	○	○	○	●	●	●	○
$w_4 = 5$	○	○	○	○	○	●	●	●	○	○	○	○	○	○	○
$-w_4 = -5$	○	○	○	○	○	○	○	○	○	○	●	●	●	○	○
$w_5 = 7$	○	○	○	○	○	○	○	●	●	●	○	○	○	○	○
$-w_5 = -7$	○	○	○	○	○	○	○	○	●	●	●	○	○	○	○
$\|N(u) \cap C\|$	5	5	5	2	2	2	2	2	2	2	2	2	2	2	2

Table 4.3: Illustration of the Theorem 4.33 to construct a $(1, 5, 2)$-code of $\mathbb{Z}^6$.

In the two-dimensional grid, i.e., the usual infinite grid, Puzynina studied the periodicity of (r, a, b)-codes. For $r = 1$, there exist non-periodic (r, a, b)-codes but, all of them can be obtained from periodic ones [Puz04]. That is to say, if r, a, b are such that there exists an (r, a, b)-code, then there exists a periodic colouring that is an (r, a, b)-code. Moreover when $r \geq 2$, the author proved that every (r, a, b)-code is periodic [Puz08].

Theorem 4.35. [Puz08]

For $r \geq 2$, every (r, a, b)-code of $\mathbb{Z}^2$ is periodic.

When the difference between a and b is large enough, the precise type of the periodic colouring is known.

Theorem 4.36. [Axe03]

If a colouring is an (r, a, b)-code of $\mathbb{Z}^2$ with $r \geq 2$ and $|a - b| > 4$, then it is one of the following diagonal colourings 1–5:

1. q-periodic colouring where $q \in \{r, r + 1\}$ is odd and the monochromatic diagonals are parallel.

2. q-anti-periodic colouring where $q \in \{r, r + 1\}$ is even.

3. q-periodic colouring where $q \in \{r, r + 1\}$ is even and for all horizontal or vertical interval I of length p the number of black vertices from the even sublattice and from the odd sublattice is the same.

4. $(2r+1)$-periodic colouring and for all horizontal or vertical interval I of length p the number of black vertices from the even sublattice and from the odd sublattice is the same.

5. 2-periodic or 3-periodic colouring.

This theorem is used in Chapter 6 to obtain the precise values of a and b for any (r, a, b)-code of $\mathbb{Z}^2$ with $r \geq 2$ and $|a - b| > 4$.

Chapter 5

Identifying codes in vertex-transitive graphs

We consider the problem of computing identifying codes of graphs and its fractional relaxation. The ratio between the size of optimal integer and fractional solutions is between 1 and $2\ln(|V|)+1$ where V is the set of vertices of the graph. We focus on vertex-transitive graphs for which we can compute the exact fractional solution. There are known examples of vertex-transitive graphs that reach both bounds. We exhibit infinite families of vertex-transitive graphs with integer and fractional identifying codes of order $|V|^{\alpha}$ with $\alpha \in \{\frac{1}{4}, \frac{1}{3}, \frac{2}{5}\}$. These families are generalized quadrangles (strongly regular graphs based on finite geometries). They also provide examples for metric dimension of graphs. This chapter is based on a joint work with my co-advisor Gravier, a postdoctoral fellow Parreau and two specialists in generalized quadrangles, Professor Storme and one of his PhD candidate Rottey.

Contents

The problem of computing an identifying code of minimal size is NP-complete in general [CHL03] but can be naturally expressed as an integer linear program. A possible way to tackle this problem is to consider the fractional relaxation of the program. Naturally, one can ask how good the fractional relaxation can be. We focus on vertex-transitive graphs since for these graphs, we are able to compute the optimal size of a fractional identifying code. This value depends only on three parameters of the graph: the number and degree of vertices and the smallest size of the symmetric difference of two distinct closed neighbourhoods. Moreover, the optimal cardinality of an integer identifying code is at most at a logarithmic factor (in the number of vertices $|V|$) of the fractional optimal value.

Identifying codes have already been studied in different classes of vertex-transitive graphs, especially in cycles [BCHL04, GMS06, JL12, XTH08] and hypercubes [BHL00, EJLR08, ELR08, HL02, KCL98]. In these examples, the order of the size of an optimal identifying code seems to always match its fractional value. However, the smallest size of symmetric differences of closed neighbourhoods is small compared to the number of vertices: either it is constant (for cycles) or it has logarithmic order in the number of vertices (for hypercubes). Therefore we focus in this chapter on strongly regular vertex-transitive graphs that are graphs with the property that two adjacent (respectively non-adjacent) vertices always have the same number of common neighbours. In particular, the size of symmetric differences can only take two values and is of order at least $\sqrt{|V|}$ if the graph is not a trivial strongly regular graph.

Another interest of considering identifying codes in strongly regular graphs is that they are strongly related to resolving sets (Corollary 4.30). In particular, the optimal size of identifying codes and the metric dimension have the same order for strongly regular graphs. Actually, resolving sets were introduced by Babai [Bab80] in order to improve the complexity of the isomorphism problem for strongly regular graphs. He established an upper bound of order $\sqrt{|V|}\log_2(|V|)$ on the metric dimension of strongly regular graphs [Bab80, Bab81]. Later, Fijavž and Mohar exhibited a family of strongly regular graphs with logarithmic metric dimension, namely Paley graphs [FM04]. Bailey and Cameron proved that the metric dimension of some Kneser and Johnson graphs has order $\sqrt{|V|}$ [BC11]. Values for small strongly regular graphs have been computed [Bai13a, KCČ$^+$08]. Recently, Bailey [Bai13b] used resolving sets in strongly regular graphs to compute the metric dimension of some distance-regular graphs (graphs for which there is an automorphism between any two pairs of vertices at the same distance).

Paley graphs give an example of an infinite family of graphs for which the optimal value of fractional identifying code is constant but the integer value is logarithmic, and so the gap between the two is also logarithmic. We consider another family of strongly regular graphs that have never been studied in the context of identifying codes nor resolving sets: the adjacency graphs of generalized quadrangles. These graphs are constructed using finite geometries. Constructing identifying codes can be seen as a way to break the inherent symmetry of these graphs. We give constructions of identifying codes with size of optimal order. This order is of the form $|V|^\alpha$ with $\alpha \in \{\frac{1}{4}, \frac{1}{3}, \frac{2}{5}\}$ and corresponds to the order of the fractional value.

This chapter is based on a joint work [GPR$^+$]. At the origin, with my co-advisor Gravier and a post-doctoral fellow at the University of Liège Parreau, we looked at the fractional relaxation of the identifying code problem and discuss the cases when the separation condition or the domination condition prevails. A poster presentation at a PhD-day organized by the Belgian Mathematical Society [GPV13] aroused the interest of Professor Storme from Gand university. As specialists in generalized quadrangles, his doctoral student and himself joined us to study identifying codes in these graphs. The collaboration was fruitful as we found constructions of identifying codes of optimal size in this setting.

This chapter is organized as follows. In Section 5.1, we exhibit the linear program for identifying codes, compute the optimal value of the relaxation for vertex-transitive graphs and deduce a general bound. In Section 5.2, we review known results for identifying codes in vertex-transitive graphs and compare them to our general bound. Finally in Section 5.3, we study strongly regular graphs and in particular adjacency graphs of generalized quadrangles.

5.1 Fractional relaxation

The problem of finding a minimal identifying code in a graph G can be expressed as a linear program P_G (Remark 4.26). Let us denote by P_G^* the linear programming fractional relaxation of P_G where the integrality condition $x_u \in \{0,1\}$ is replaced by a linear constraint $0 \le x_u \le 1$ for all vertices $u \in V$.

$$\begin{array}{llllr}
\text{Minimize} & \displaystyle\sum_{x_u \in V} x_u & & & \\
\text{such that} & \displaystyle\sum_{w \in N[u]} x_w & \ge 1 & \forall u \in V & \text{(domination)} \\
 & \displaystyle\sum_{w \in N[u]\Delta N[v]} x_w & \ge 1 & \forall u, v \in V, u \neq v & \text{(separation)} \\
 & x_u \in [0,1] & & \forall u \in V &
\end{array}$$

The optimal value of P_G^*, denoted by $\gamma_f^{\mathrm{ID}}(G)$, gives an estimation on $\gamma^{\mathrm{ID}}(G)$ within a logarithmic factor.

Proposition 5.1.

Let $G = (V, E)$ be a twin-free graph. We have

$$\gamma_f^{\mathrm{ID}}(G) \le \gamma^{\mathrm{ID}}(G) \le \gamma_f^{\mathrm{ID}}(G)(1 + 2\ln|V|).$$

Proof. The first inequality is trivial since P_G^* is a relaxation of P_G. Recall from Remark 4.22, that the identifying code problem in $G = (V, E)$ is equivalent to a covering problem in the hypergraph $\mathcal{H} = (V, \mathcal{E})$ with hyperedge set

$$\mathcal{E} = \{N[u] \mid u \in V\} \cup \{N[u]\Delta N[v] \mid u \neq v \in V\}.$$

The covering problem in $\mathcal{H}$ is to find a set of vertices of minimum size that intersects all the hyperedges. The linear programming formulations of the two problems are the same. Using the result of Lovász [Lov75] on the ratio of optimal integral and fractional covers, we have

$$\gamma^{\mathrm{ID}}(G) \le \gamma_f^{\mathrm{ID}}(G)(1 + \ln r)$$

where r is the maximal degree of $\mathcal{H}$, i.e., the maximal number of hyperedges a vertex is belonging to. Let $u \in V$ and k be its degree in G. Then u is in $k+1$ hyperedges of the form

$N[v]$ and in $(|V|-k-1)(k+1)$ hyperedges of the form $N[v]\Delta N[w]$. Indeed, we must have $v \in N[u]$ and $w \notin N[u]$. Hence the degree of u in $\mathcal{H}$ is $(|V|-k)(k+1)$. The maximal value of $(|V|-k)(k+1)$ with $0 \leq k \leq |V|-1$ is obtained for $k = \frac{|V|-1}{2}$. Therefore, $r \leq \frac{(|V|+1)^2}{4} \leq |V|^2$ for $|V| \neq 0$ which leads to the upper bound of the proposition. □

In the case of vertex-transitive graphs, we can compute the exact value of γ_f^{ID}.

Proposition 5.2.

Let $G = (V, E)$ be a twin-free vertex-transitive graph. Let k denote the degree of G and let d denote the smallest size of symmetric differences of closed neighbourhoods $N[u]\Delta N[v]$ among all the pairs of distinct vertices u, v of V. We have

$$\gamma_f^{\mathrm{ID}}(G) = \frac{|V|}{\min(k+1, d)}.$$

In particular

$$\frac{|V|}{\min(k+1, d)} \leq \gamma^{\mathrm{ID}}(G) \leq \frac{|V|(1 + 2\ln|V|)}{\min(k+1, d)}.$$

Proof. Giving to each variable x_u the value $\frac{1}{\min(k+1,d)}$ leads to a feasible solution of P_G^*, hence

$$\gamma_f^{\mathrm{ID}}(G) \leq \frac{|V|}{\min(k+1, d)}.$$

Since G is a vertex-transitive graph, all the vertices play the same role. Consider the finite set $\mathcal{S}$ of extreme optimal solutions (solutions that are vertices of the polytope defined by P_G^*). Any linear combination of elements of $\mathcal{S}$, with the sum of coefficients equal to 1, is still an optimal solution of P_G^*. In particular, $\mathbf{x} = \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}} s$ is an optimal solution. We claim that all the components of $\mathbf{x}$ are equal. Indeed, assume that $x_u \neq x_v$ and let μ be an automorphism sending u to v. Let $s \in \mathcal{S}$, then $\mu(s)$ and $\mu^{-1}(s)$, obtaining by permuting the value inside s following the automorphism μ are still extreme optimal solutions. Hence $\mathcal{S}$ is stable by μ and so $\mu(\mathbf{x}) = \mathbf{x}$, a contradiction since $x_u \neq x_v$. □

Remark 5.3. Let $G = (V, E)$ be a vertex-transitive graph with degree k and let d denote the smallest size of symmetric differences between closed neighbourhoods. If $d < k+1$, then $\gamma_f^{\mathrm{ID}}(G) = \frac{n}{d}$ from the previous proposition. In this case, we say that *the separation condition prevails.* If $k+1 < d$, then $\gamma_f^{\mathrm{ID}}(G) = \frac{n}{k+1}$ and we say that *the domination condition prevails.*

5.2 Known results on transitive graphs

We review some known results on classes of transitive graphs. In particular, we discuss the gap between γ^{ID} and γ_f^{ID}. Sometimes, not only identifying codes but also *r-identifying codes* have been studied in these classes. Instead of using the closed neighbourhoods, that are the balls of radius 1, one consider the balls of radius r to identify the vertices. It is equivalent to consider r-identifying codes in a graph G or to consider identifying codes in G^r, the r^{th}-power of G, obtained by adding edges between each pair of vertices of G that are at distance at most r. In the following, we express the results in terms of identifying codes in the power graph.

5.2.1 Cycles

We first consider cycles and powers of cycles. Let $n, r \in \mathbb{N}$ with $n \geq 5$ and $1 \leq r < \frac{n-1}{2}$. Recall from Subsection 4.3.1 that the cycle $\mathcal{C}_n$ on n vertices has vertex set $V = \{0, 1, ..., n-1\}$ and two distinct vertices i and j are adjacent if $|i-j| = 1$ (modulo n). The graph $\mathcal{C}_n^r$ is still vertex-transitive with vertex degree $2r$. The smallest symmetric difference of closed neighbourhoods has size 2. It is obtained via two consecutive vertices i and $i+1$ whose symmetric difference of closed neighbourhoods is the set $\{i-r, i+r+1\}$ (modulo n). Hence the fractional identifying code value is $\gamma_f^{\mathrm{ID}}(\mathcal{C}_n^r) = \frac{n}{2}$.

On the other hand, the study of integer identifying codes in power of cycles had taken several years (see *e.g.* [BCHL04, GMS06, XTH08]) before being completed by Junnila and Laihonen [JL12]. We have the following results. If n is even and at least $2r+4$, then

$$\gamma^{\mathrm{ID}}(\mathcal{C}_n^r) = \frac{n}{2} = \gamma_f^{\mathrm{ID}}(\mathcal{C}_n^r).$$

If n is odd and at least $2r+3$, then

$$\frac{n+1}{2} \leq \gamma^{\mathrm{ID}}(\mathcal{C}_n^r) \leq \frac{n+1}{2} + r.$$

In particular, the difference between $\gamma^{\mathrm{ID}}(\mathcal{C}_n^r)$ and $\gamma_f^{\mathrm{ID}}(\mathcal{C}_n^r)$ is bounded by r . Hence the ratio is converging to 1 when r is fixed and n is large.

When $n = 2r+2$, $\mathcal{C}_n^r$ is a complete graph where a perfect matching is removed. See for instance Figure 5.1. In this case, we have $\gamma^{\mathrm{ID}}(\mathcal{C}_n^r) = n-1$. Then $\frac{\gamma^{\mathrm{ID}}(\mathcal{C}_n^r)}{\gamma_f^{\mathrm{ID}}(\mathcal{C}_n^r)} \to 2$ when n is large. Finally, if $n = 2r+3$, $\gamma^{\mathrm{ID}}(\mathcal{C}_n^r) = \lfloor \frac{2n}{3} \rfloor$ and the ratio is converging to 4/3.

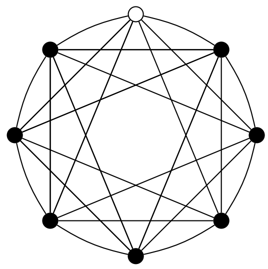


Figure 5.1: The set of black vertices is an identifying code of $\mathcal{C}_8^3$ of minimal size.

5.2.2 Hypercubes

Let $q \geq 3$. Recall from Subsection 4.3.2, that each vertex of the hypercube $\mathcal{H}_q$ has degree q. The smallest symmetric difference of closed neighbourhoods in $\mathcal{H}_q$ has size $d = 2q-2$ and is obtained via two adjacent vertices. Hence, by Proposition 5.2,

$$\gamma_f^{\mathrm{ID}}(\mathcal{H}_q) = \frac{2^q}{q+1}.$$

Computing the exact value of $\gamma^{\mathrm{ID}}(\mathcal{H}_q)$ seems difficult and only few exact values are known. However, we have the following bounds (see [EJLR08, Theorem 4] for the upper bound and [KCL98] for the lower bound)

$$\frac{q2^{q+1}}{q(q+1)+2} \leq \gamma^{\mathrm{ID}}(\mathcal{H}_q) \leq \frac{9}{2} \cdot \frac{2^q}{q+1}.$$

Hence the integer and the fractional identifying code values have the same order and the ratio satisfies

$$2 - \frac{4}{q(q+1)+2} \leq \frac{\gamma^{\mathrm{ID}}(\mathcal{H}_q)}{\gamma_f^{\mathrm{ID}}(\mathcal{H}_q)} \leq \frac{9}{2}.$$

Let $1 < r < q$. We now consider r-identifying codes or equivalently identifying codes in $\mathcal{H}_q^r$. The graph $\mathcal{H}_q^r$ is still vertex-transitive. The degree of the vertices is $k = \sum_{i=1}^{r} \binom{q}{i}$. The smallest symmetric difference of closed neighbourhoods has now size $d = 2\binom{q-1}{r}$ and is still done by two adjacent vertices of $\mathcal{H}_q$. Thus, by Proposition 5.2,

$$\gamma_f^{\mathrm{ID}}(\mathcal{H}_q^r) = \frac{2^q}{\min\left(\sum_{i=0}^{r} \binom{q}{i}, 2\binom{q-1}{r}\right)}.$$

Concerning the general behaviour of $\gamma^{\mathrm{ID}}(\mathcal{H}_q^r)$, we consider two cases: r is fixed or r is linearly dependent of q. Assume first that r is fixed and q is large. The bounds given by Karpovsky et al. [KCL98] can be translated as follows. There are two constants α and β (depending on r) such that, for large q,

$$\alpha \frac{2^q}{q^r} \leq \gamma^{\mathrm{ID}}(\mathcal{H}_q^r) \leq \beta \frac{2^q}{q^r}. \tag{5.1}$$

Thus $\gamma^{\mathrm{ID}}(\mathcal{H}_q^r)$ and $\gamma_f^{\mathrm{ID}}(\mathcal{H}_q^r)$ have the same order, that is $2^q/q^r$.

Assume now that $r = \lfloor \rho q \rfloor$ for some constant ρ. Honkala and Lobstein [HL02] proved that

$$\lim_{q \to \infty} \frac{\log_2 \gamma^{\mathrm{ID}}(\mathcal{H}_q^r)}{q} = 1 - h(\rho)$$

where $h(x) = -x \log_2(x) - (1-x)\log_2(1-x)$ is the binary entropy function. This result can be proved with Proposition 5.1. Indeed, $\frac{\log_2 \sum_{i=0}^{r} \binom{q}{i}}{q}$ and $\frac{\log_2 \binom{q}{r}}{q}$ tend to $h(\rho)$. Hence

$$\lim_{q \to \infty} \frac{\log_2 \gamma_f^{\mathrm{ID}}(\mathcal{H}_q^r)}{q} = 1 - h(\rho)$$

and

$$\lim_{q \to \infty} \frac{\log_2 \left(\gamma_f^{\mathrm{ID}}(\mathcal{H}_q^r)(1 + 2\ln 2^q)\right)}{q} = 1 - h(\rho).$$

But we do not know if $\gamma^{\mathrm{ID}}(\mathcal{H}_q^r)$ and $\gamma_f^{\mathrm{ID}}(\mathcal{H}_q^r)$ have the same order in this case.

5.2.3 Product of graphs

One can easily obtain other vertex-transitive graphs by doing products of vertex-transitive graphs such as cliques[1]. Identifying codes in the following products of graphs have been recently considered.

[1] This was already the case for hypercubes which are Cartesian products of q cliques of size 2.

Cartesian product of two cliques. Let $2 \leq p \leq q$ be integers. The Cartesian product $K_p \square K_q$ of two cliques is a vertex-transitive graph with vertex degree $k = p + q - 1$. The smallest symmetric difference of closed neighbourhoods has sized $d = 2p - 2$ and is obtained via two adjacent vertices. By Proposition 5.2,

$$\gamma_f^{\mathrm{ID}}(K_p \square K_q) = \frac{pq}{2p-2}.$$

Identifying codes in $K_p \square K_q$ have been studied by Gravier et al. [GMS08] when the two cliques are of the same size, and by Goddard and Wash [GW13] in the general case. When $q = p$, Gravier et al. [GMS08] proved that $\gamma^{\mathrm{ID}}(K_p \square K_p) = \lfloor \frac{3p}{2} \rfloor$ and an identifying code of $K_p \square K_p$ of minimal size is given by the set of vertices of a diagonal of the graph together with half of the vertices of the other diagonal, as depicted in Figure 5.2. Goddard and Wash [GW13] proved that

$$\gamma^{\mathrm{ID}}(K_p \square K_q) = \begin{cases} q + \lfloor \frac{p}{2} \rfloor & \text{if } q \leq \frac{3p}{2} \\ 2q - p & \text{if } q \geq \frac{3p}{2}. \end{cases}$$

Therefore, the ratio between integer and fractional identifying codes values is

$$\frac{\gamma^{\mathrm{ID}}(K_p \square K_q)}{\gamma_f^{\mathrm{ID}}(K_p \square K_q)} = \begin{cases} 2 + \frac{p}{q} - \frac{2}{p} - \frac{1}{q} & \text{if } q \leq \frac{3p}{2} \\ 4 - \frac{2p}{q} - \frac{4}{p} + \frac{2}{q} & \text{if } q \geq \frac{3p}{2} \end{cases}$$

In particular, it is bounded by a constant.

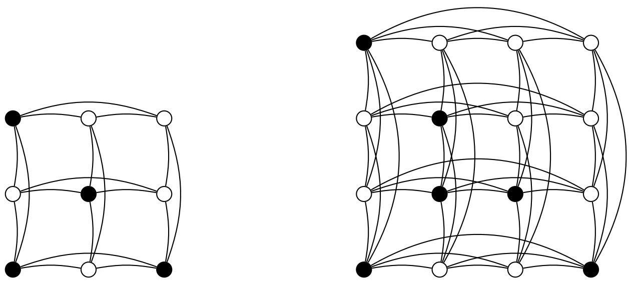


Figure 5.2: The set of black vertices is an identifying code of $K_p \square K_p$ of minimal size, with $p = 3$ and $p = 4$.

Direct product of cliques Let $2 \leq p \leq q$ be integers. The direct product $K_p \times K_q$ of two cliques is a vertex-transitive graph with vertex degree $k = (p-1)(q-1)$. The smallest symmetric difference of closed neighbourhoods has sized $d = 2p$ and is obtained via two vertices belonging to the same copy of K_q. By Proposition 5.2,

$$\gamma_f^{\mathrm{ID}}(K_p \times K_q) = \begin{cases} \frac{q}{2} & \text{if } p \geq 4 \text{ or } q > p \\ \frac{pq}{(p-1)^2+1} & \text{if } p \leq 3 \text{ and } p = q. \end{cases}$$

Rall and Wash [RW14] gave the exact size of optimal identifying codes in $K_p \times K_q$. Except the small values of p and q, there are two main cases. If $p \geq 3$ and $q \geq 2p$, then $\gamma^{\mathrm{ID}}(K_p \times K_q) = q-1$. If $p \geq 5$ and $q < 2p$, then $\gamma^{\mathrm{ID}}(K_p \times K_q)$ is either $\lfloor \frac{2(p+q)}{3} \rfloor$ or $\lceil \frac{2(p+q)}{3} \rceil$ depending on the value of $p+q$ modulo 3. Therefore, the ratio between integer and fractional identifying codes values is either $2 - 2/q$ or $4/3(1+p/q)$ and is again bounded.

Lexicographic product of graphs. Let G and H be two vertex-transitive graphs that are not complete graphs. Then $G[H]$ is also vertex-transitive. If G (respectively H) has vertex degree k_G (resp. k_H) and n_G (resp. n_H) vertices, then $G[H]$ has $n_G n_H$ vertices and vertex degree $k = k_G n_H + k_H$. Moreover, the size of the smallest symmetric difference of closed neighbourhoods of $G[H]$ and H are equal. Hence

$$\gamma_f^{\mathrm{ID}}(G[H]) = \frac{n_G n_H}{d_H}$$

where d_H is the smallest symmetric difference of closed neighbourhoods of H.

Assume that G does not have two vertices u and v such that $N(u) = N(v)$. Feng et al. [FXW12] proved that in this case

$$\gamma^{\mathrm{ID}}(G[H]) = n_G s_H$$

where s_H is the minimum size of a separating set of H. Hence we have

$$\frac{\gamma^{\mathrm{ID}}(G[H])}{\gamma_f^{\mathrm{ID}}(G[H])} = \frac{s_H d_H}{n_H}.$$

If H is such that $k_H + 1 \geq d_H$, then $\gamma_f^{\mathrm{ID}}(H) = \frac{n_H}{d_H}$. Since s_H is either equal to $\gamma^{\mathrm{ID}}(H)$ or $\gamma^{\mathrm{ID}}(H) - 1$, the ratio between $\gamma^{\mathrm{ID}}(G[H])$ and $\gamma_f^{\mathrm{ID}}(G[H])$ is the same than the one for H. In particular, if we have a ratio α for a graph H we can get graphs with arbitrary sizes and still ratio α.

5.3 Strongly regular graphs

The bound of Proposition 5.2 is helpful when the symmetric differences are large (larger than $\ln |V|$). For this reason, we now focus on the family of strongly regular graphs for which the smallest symmetric difference has, in most cases, size at least $\sqrt{|V|}$ (see Proposition 4.18). Recall from Subsection 4.3.3, that a strongly regular graph $\mathrm{srg}(n, k, \lambda, \mu)$ is a k-regular graph G on n vertices for which any pair of adjacent (respectively non-adjacent) vertices have exactly λ (resp. μ) neighbours in common.

Strongly regular graphs have been used once for the problem of identifying codes by Gravier et al. [GJLR14] to provide families of graphs for which all the subsets of a given size are identifying codes. However, they did not study optimal identifying codes. On the opposite and as mentioned in the introduction, resolving sets and metric dimension have been studied in several contexts for strongly regular graphs. In particular, Babai [Bab80] gave an upper bound on the size of the symmetric differences of open neighbourhood in strongly regular graphs which lead to bounds on the metric dimension. Following these ideas, we prove similar results for identifying codes.

We first compute the smallest size d of the symmetric differences of closed neighbourhoods using λ and μ and then give a general upper bound on d.

Proposition 5.4.

Let G be a strongly regular graph $\mathrm{srg}(n, k, \lambda, \mu)$. Let u and v be two vertices of G. If u is adjacent to v, then $|N[u]\Delta N[v]| = 2(k-1) - 2\lambda$. Otherwise, we have $|N[u]\Delta N[v]| = 2(k+1) - 2\mu$. Hence, the smallest symmetric difference of closed neighbourhoods is

$$d = \min(2(k-\lambda-1), 2(k-\mu+1)) = 2k - 2\max(\lambda+1, \mu-1).$$

If G is vertex-transitive, we have

$$\gamma_f^{\mathrm{ID}}(G) = \frac{n}{\min(k+1, 2(k-\lambda-1), 2(k-\mu+1))}.$$

Proof. Let u and v be two adjacent vertices. There are $k-\lambda$ neighbours of u that are not neighbours of v. But v is counted in these vertices. Hence $|N[u] \setminus N[v]| = k-1-\lambda$ and we get the results. The computation for the non-adjacent case is similar. □

Remark 5.5. Actually, it seems that almost all the strongly regular graphs are not-vertex transitive, see for example [KÖ04]. However, all the strongly regular graphs we are considering in the following are in fact vertex-transitive.

From Proposition 5.4 and Proposition 5.2 together with the bounds obtained in Proposition 4.18, we derive the following general bound for strongly regular graphs when they are vertex-transitive.

Corollary 5.6.

Let G be a primitive strongly regular graph $\mathrm{srg}(n, k, \lambda, \mu)$. If G is vertex-transitive, we have

$$\gamma^{\mathrm{ID}}(G) \leq \frac{n(1+2\ln n)}{\sqrt{n}-3}.$$

In particular $\gamma^{\mathrm{ID}}(G) = O(\sqrt{n}\ln n)$.

5.3.1 Known results on particular families

The only strongly regular graphs for which we know optimal identifying codes are Cartesian and direct product of two cliques of the same size that we already mentioned in the previous section. The Cartesian product $K_p \square K_p$ is a strongly regular graph $\mathrm{srg}(p^2, 2p-2, p-2, 2)$ whereas $K_p \times K_p$ (that is the complement of $K_p \square K_p$) is a $\mathrm{srg}(p^2, (p-1)^2, (p-2)^2, (p-2)(p-1))$. We obtain results for some other families by considering the previous work on metric dimension.

Kneser and Johnson graphs (of diameter 2). Let $1 \leq p \leq m$. The *Johnson graph* $J(m, p)$ is the graph whose vertices are the subsets of size p of a set of m elements and two vertices are adjacent if the corresponding sets intersect in exactly $p-1$ elements. Since the diameter of $J(m,p)$ is $\min(p, m-p)$, the graph $J(m,p)$ is a primitive strongly regular graph if and only if $p=2$ or $p=m-2$. Note that the two corresponding graphs are isomorphic and have parameters $\mathrm{srg}(\binom{m}{2}, 2(m-2), m-2, 4)$.

The *Kneser graph* $K(m,p)$ is the graph whose vertices are the subsets of size p of a set of m elements and two vertices are adjacent if the corresponding sets do not intersect. The Kneser graph $K(5,2)$ corresponds to the well known Petersen graph (Figure 5.3). The graph $K(m,p)$ is a primitive strongly regular graph if and only if $p=2$ and $K(m,2)$ is a $\mathrm{srg}(\binom{m}{2}, \binom{m-2}{2}, \binom{m-4}{2}, \binom{m-3}{2})$.

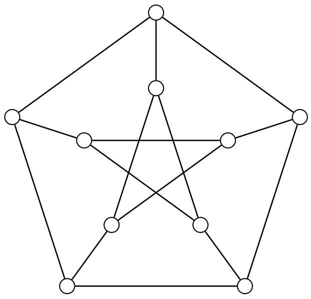


Figure 5.3: The Kneser graph $K(5,2)$, also known as the Petersen graph.

Bailey and Cameron [BC11] have computed the exact value of the metric dimension in $J(m,2)$ and $K(m,2)$.

Proposition 5.7. [BC11, Corollary 3.33]

For $m \geq 6$, the metric dimension of the Johnson graph $J(m,2)$ and the Kneser graph $K(m,2)$ is $\frac{2}{3}(m-i)+i$ where $m \equiv i \bmod 3$.

Using Corollary 4.30 we obtain a bound for identifying codes.

Corollary 5.8.

Let G be $K(m,2)$ or $J(m,2)$. We have

$$\frac{2m}{3} \leq \gamma^{\mathrm{ID}}(G) \leq \frac{4(m+1)}{3}.$$

In particular, $\gamma^{\mathrm{ID}}(G) \in \Theta(\sqrt{|V|})$.

To compute the fractional identifying code number, one just has to compute the value of the smallest symmetric difference using Proposition 5.4. For $K(m,2)$ and $m \geq 6$, $\mu-1 \geq \lambda+1$ and $2k-2\mu+2 = 2(m-1) \leq k+1$. Hence, for $m \geq 6$,

$$\gamma_f^{\mathrm{ID}}(K(m,2)) = \frac{m(m-1)}{4(m-1)} = \frac{m}{4}.$$

For $J(m,2)$, $\lambda+1 \geq \mu-1$ whenever $m \geq 4$ and $2k-2\lambda-2 = 2(m-3) = k$. Hence

$$\gamma_f^{\mathrm{ID}}(J(m,2)) = \frac{m(m-1)}{4(m-3)} = \frac{m}{4} + 2.$$

In all cases, we have $\gamma_f^{\mathrm{ID}}(G) \in \Theta(\sqrt{|V|})$ and the fractional and integer values have the same order for these graphs.

Paley graphs. The Paley graph P_q is defined for a prime power $q = 1 \bmod 4$. Vertices are the elements of the finite field $\mathbb{F}_q$ on q elements, and a is adjacent to b if $a-b$ is a square. They are strongly regular srg $\left(q, \frac{1}{2}(q-1), \frac{1}{4}(q-5), \frac{1}{4}(q-1)\right)$. Paley graphs have the particularity to have symmetric difference of closed neighbourhoods of order $|V|$, hence the fractional identifying code number is bounded by a constant and the identifying code number is of order $\log_2 |V|$.

Proposition 5.9.

Let q be a prime power satisfying $q = 1 \bmod 4$ and $q \geq 9$. We have $\gamma_f^{\mathrm{ID}}(P_q) = \frac{2q}{q-1}$ and thus

$$\log_2(q+1) \leq \gamma^{\mathrm{ID}}(P_q) \leq (2+o(1))(1+2\ln q).$$

In particular, $\gamma^{\mathrm{ID}}(P_q) \in \Theta(\log_2 |V|)$.

Proof. We first compute the value of d. We have

$$\max(\lambda+1, \mu-1) = \tfrac{q-1}{4}.$$

Thus $d = \frac{q-1}{2} < k+1 = \frac{q+1}{2}$ and $\gamma_f^{\mathrm{ID}}(P_q) = \frac{2q}{q-1} \leq 2 + o(1)$.

The lower bound on $\gamma^{\mathrm{ID}}(P_q)$ is the general lower bound of Proposition 4.23. For the upper bound, we use the bound of Proposition 5.1 with $\gamma_f^{\mathrm{ID}}(P_q)$. □

Similar results were obtained for metric dimension.

Proposition 5.10. Fijavž and Mohar [FM04]

Let q be a prime power satisfying $q = 1 \bmod 4$. Then the metric dimension of the Paley graph P_q satisfies

$$\log_2 q \leq \beta(P_q) \leq 2\log_2 q.$$

In particular, $\beta(P_q) \in \Theta(\log_2 |V|)$.

5.4 Generalized quadrangles

The graphs obtained from generalized quadrangles form another family of strongly regular graphs. This family provides examples of vertex-transitive graphs where the domination condition prevails for the value of γ_f^{ID}. Let s, t be positive integers. A *generalized quadrangle* $\mathrm{GQ}(s,t)$ is an incidence structure, i.e., a set of points and lines, such that

- there are $s+1$ points on each line,
- there are $t+1$ lines passing through each point,
- for a point P that does not lie on a line L, there is exactly one line passing through P and intersecting L.

Such an incidence structure has $(st+1)(s+1)$ points and $(st+1)(t+1)$ lines. A trivial example is the incidence structure given by a square grid of size $s \times s$ which is a $\mathrm{GQ}(s-1,1)$. The *dual* of a generalized quadrangle is obtained by reversing the role of the lines and the points.

Adjacency graphs can be naturally obtained from generalized quadrangles: consider the points as vertices and two vertices are adjacent if the corresponding points belong to a common line. For example, a bipartite complete graph $K_{t,t}$ is a $\mathrm{GQ}(1,t-1)$. By abuse of notation, $\mathrm{GQ}(s,t)$ will also denote the adjacency graph of a generalized quadrangle with parameters s and t.

Observe that a $\mathrm{GQ}(s,t)$ is a strongly regular graph

$$\mathrm{srg}((st+1)(s+1), s(t+1), s-1, t+1).$$

Indeed, any vertex has degree $k = s(t+1)$, any pair of adjacent vertices has $s-1$ common neighbours and any pair of non-adjacent vertices has $t+1$ common neighbours. From these values, we can easily compute the smallest size of symmetric differences of closed neighbourhoods: $d = 2s(t+1) - 2\max(s,t)$. We have $d > k+1$ if and only if $\mathrm{GQ}(s,t)$ is not trivial, i.e., $s > 1$ and $t > 1$. In other words, In that case, the following inequalities, for which Cameron gave a short combinatorial proof [Cam75], hold.

Lemma 5.11. Higman's inequality [Hig71, Hig74]

For a $\mathrm{GQ}(s,t)$, if $s > 1$ and $t > 1$, then $t \leq s^2$ and dually $s \leq t^2$.

From now on, we assume that $s > 1$ and $t > 1$. We obtain the following bounds on γ_f^{ID} for generalized quadrangles.

Proposition 5.12.

Let G be a vertex-transitive $\mathrm{GQ}(s,t)$ with $s > 1$ and $t > 1$. Let n denote the number of vertices of the graph G. We have

$$2^{-5/4} \cdot n^{1/4} \leq \gamma_f^{\mathrm{ID}}(G) \leq 2 \cdot n^{2/5}.$$

Proof. Let G be a vertex-transitive adjacency graph of a $\mathrm{GQ}(s,t)$ with $s > 1$ and $t > 1$. Then $n = (st+1)(s+1)$ is the number of vertices of G. We have by Proposition 5.2

$$\gamma_f^{\mathrm{ID}}(G) = \frac{(st+1)(s+1)}{s(t+1)+1} = \frac{s^2 t}{st+s+1} + 1.$$

As $st < st + s + 1 < 2st$, we obtain $\frac{1}{2}s < \gamma_f^{\mathrm{ID}}(G) < 2s$.

Moreover, using the previous lemma, we get

$$s^{5/2} \le s^2 t < s^2 t + st + s + 1 = n \le s^4 + s^3 + s + 1 < 2 \cdot s^4.$$

So $\left(\frac{1}{2}n\right)^{1/4} < s < n^{2/5}$. It follows that $\left(\frac{1}{2}\right)^{5/4} \cdot n^{1/4} < \gamma_f^{\mathrm{ID}}(G) < 2 \cdot n^{2/5}$. □

Constructions of GQ(s,t) are known only for (s,t) or (t,s) in the set

$$\{(q,q),(q,q^2),(q^2,q^3),(q-1,q+1)\}$$

where q is a prime power. Many of them are based on finite geometries. Generalized quadrangles coming from finite classical polar spaces of rank 2 are given in Table 5.1. For more information on these geometric structures, see e.g. [HT91]. It is well known that these polar spaces give rise to generalized quadrangles and they are often referred to as the *classical generalized quadrangles* [PT84].

Polar space	Name	(s,t)	
$Q^+(3,q)$	Hyperbolic	$(q,1)$	a grid
$Q(4,q)$	Parabolic	(q,q)	dual of $W(3,q)$
$Q^-(5,q)$	Elliptic	(q,q^2)	dual of $H(3,q^2)$
$H(3,q^2)$	Hermitian	(q^2,q)	dual of $Q^-(5,q)$
$H(4,q^2)$	Hermitian	(q^2,q^3)	
$W(3,q)$	Symplectic	(q,q)	dual of $Q(4,q)$

Table 5.1: The finite classical polar spaces of rank 2.

Example 5.13 (The grid GQ$(q,1)$)**.** Let q be a prime power. We set ourselves in the 3-dimensional projective space PG$(3,q)$ over the finite field $\mathbb{F}_q$. The points of PG$(3,q)$ can be described using four coordinates $(X_0,X_1,X_2,X_3) \in \mathbb{F}_q^4 \setminus \{(0,0,0,0)\}$ where two coordinates that are proportional refer to the same point.

Let Q be the set of points of PG$(3,q)$ that satisfy the equation $X_0X_1 + X_2X_3 = 0$ (Q is a hyperbolic quadric). The incidence structure $Q^+(3,q)$ obtained from the $(q+1)^2$ points of Q and $2(q+1)$ lines of Q (i.e., lines of PG$(3,q)$ included in Q) is a generalized quadrangle GQ$(q,1)$. Any point of this quadric is of the form $(bd,ac,ad,bc) \neq (0,0,0,0)$, $a,b,c,d \in \mathbb{F}_q$. There are two sets of lines on this quadric. Lines of the first type arise as the intersection of the planes $aX_0 - bX_2 = 0$ and $bX_1 - aX_3 = 0$. The $q+1$ lines of the second type are the intersection of the planes $cX_0 - dX_3 = 0$ and $dX_1 - cX_2 = 0$.

It is easy to see that $Q^+(3,q)$ is isomorphic to a grid with $q+1$ points on each line. The adjacency graph of the grid is the Cartesian product of two cliques of size $q+1$, $K_{q+1}\square K_{q+1}$. As mentioned in Section 5.2.3, the optimal cardinality of identifying codes in these graphs is $\gamma^{\mathrm{ID}}(K_s\square K_s) = \frac{3s}{2}$ [GMS08] and the fractional optimal value is $\gamma_f^{\mathrm{ID}}(K_s\square K_s) = \frac{s^2}{2(s-1)}$. Hence they have the same order.

There are other generalized quadrangles known, however they have the same parameters as one given in Table 5.1 or they have parameters $(q-1,q+1)$ or $(q+1,q-1)$. We provide identifying codes of optimal order for some cases.

5.4.1 Identifying codes in $T_2^*(\mathcal{O})$, a particular $\mathrm{GQ}(q-1, q+1)$

Proposition 5.14.

Let $q > 2$ be a power of 2. There exists a $\mathrm{GQ}(q-1, q+1)$ with an identifying code of size $3q-3 \in \Theta(n^{1/3})$ where n is the number of vertices.

Before giving the proof, we will consider a particular construction of a $\mathrm{GQ}(q-1, q+1)$ and give some structural properties.

Let q be a power of 2. We consider points of the 3-dimensional projective space $\mathrm{PG}(3, q)$ over the finite field $\mathbb{F}_q$ of order q as in Example 5.13. Consider the hyperplane H_∞ of equation $X_0 = 0$ in $\mathrm{PG}(3, q)$ and the conic $\mathcal{C}$ of equation $X_1X_3 - X_2^2 = 0$ in the hyperplane H_∞. Any line of H_∞ intersects $\mathcal{C}$ in 0, 1 or 2 points. A line intersecting $\mathcal{C}$ in one point is *tangent* to $\mathcal{C}$. There is a special point, $N(0, 0, 1, 0)$, called the *nucleus* of $\mathcal{C}$, that lies on all tangents of $\mathcal{C}$. Then any other point of H_∞ lies on exactly one tangent of $\mathcal{C}$. The set $\mathcal{O} = \mathcal{C} \cup \{N\}$ is a *hyperconic*. This set has the property that each line of H_∞ intersects $\mathcal{O}$ in 0 or 2 points.

Consider now the following incidence structure $T_2^*(\mathcal{O}) = (\mathcal{P}, \mathcal{L})$, where the set $\mathcal{P}$ of points is the set of affine points, i.e., points not in H_∞ and the set $\mathcal{L}$ of lines is the set of the lines through a point of $\mathcal{O}$ not lying in H_∞. It is well-known in geometry that the incidence structure $T_2^*(\mathcal{O})$ is a generalized quadrangle with parameters $q-1$ and $q+1$ (see [PT84, Theorem 3.1.3.]). Since this structure plays a important role in the sequel, we give a proof of this fact in this manuscript so that non-familiar readers have the opportunity to understand the incidence structure.

Theorem 5.15. [PT84]

The incidence structure $T_2^*(\mathcal{O})$ is a $\mathrm{GQ}(q-1, q+1)$.

Proof. Each line of $\mathrm{PG}(3, q)$ contains $q+1$ points and if it does not lie in H_∞, then it intersects H_∞ in exactly one point. Hence, each line of $\mathcal{L}$ has q points of $\mathcal{P}$. Since $\mathcal{O}$ contains $q+2$ points, there are $q+2$ lines of $\mathcal{L}$ going through a point of $\mathcal{P}$.

Let $\ell \in \mathcal{L}$ and $P \in \mathcal{P}$ a point not in ℓ. First, we construct a line ℓ' of $\mathcal{L}$ incident with P and intersecting ℓ in $Q \in \mathcal{P}$ (Figure 5.4(a)). Let P_∞ be the intersection point of ℓ and H_∞. By definition of the structure, $P_\infty \in \mathcal{O}$. Consider the plane π containing ℓ and P. It intersects H_∞ on a line ℓ_∞ incident with P_∞. The line ℓ_∞ intersects $\mathcal{O}$ in 0 or 2 points. Since P_∞ already belongs to the intersection, there exists another point P'_∞ of $\mathcal{O}$ that lies on ℓ_∞. Consider now the line ℓ' incident with P'_∞ and P. This line is an element of $\mathcal{L}$ and must intersect ℓ in a point Q not lying in H_∞, i.e., an element of $\mathcal{P}$.

Secondly, we prove that the projection of P on ℓ is unique (Figure 5.4(b)). Assume there is another line ℓ'' of $\mathcal{L}$ incident with P and intersecting ℓ in another point Q'. Consider again the hyperplane π containing P and ℓ. It contains also the lines ℓ' and ℓ''. So the intersection $\ell'' \cap H_\infty$ is a point P''_∞ lying on the line ℓ_∞. Hence, the points P_∞, P'_∞ and P''_∞ are three collinear points of $\mathcal{O}$, which is a contradiction. □

We will now construct an identifying code in $T_2^*(\mathcal{O})$. In $T_2^*(\mathcal{O})$, the neighbourhood of a point P is composed of a *cone* $P\mathcal{C}$ (all the lines going through P and a point of $\mathcal{C}$) and the

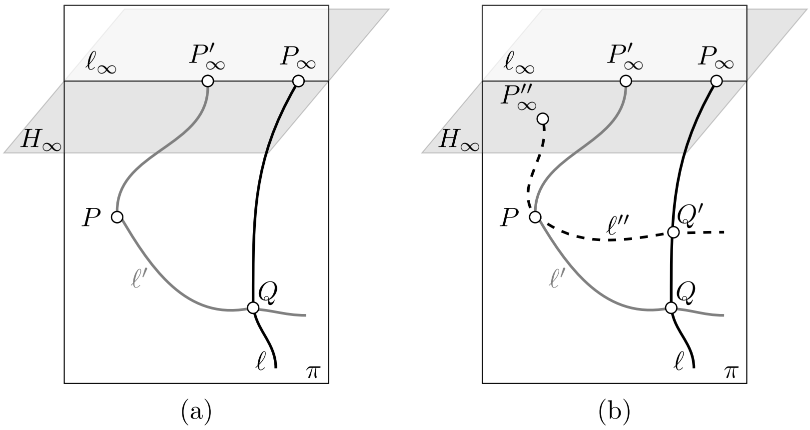


Figure 5.4: There exists a unique projection of a point on a line in $T_2^*(\mathcal{O})$.

line PN, where the points of H_∞ are removed. The common neighbours of two adjacent vertices are the $q-2$ points lying on the unique line incident with these two vertices. In the case of non-adjacent vertices, we first determine the intersection of their two cones.

Lemma 5.16.

Consider two distinct affine points P and Q such that $PQ \cap H_\infty \notin \mathcal{O}$. The intersection of the two cones $P\mathcal{C}$ and $Q\mathcal{C}$ consists of the points of the conic $\mathcal{C}$ and of points lying in a plane containing N and $PQ \cap H_\infty$.

Proof. Consider two distinct affine points $P(1,a,b,c)$ and $Q(1,\alpha,\beta,\gamma)$ such that the intersection point $PQ \cap H_\infty$ does not belong to $\mathcal{O}$. Consider the cones $P\mathcal{C}$ and $Q\mathcal{C}$ in $\mathrm{PG}(3,q)$. It is clear that the conic $\mathcal{C}$ belongs to $P\mathcal{C} \cap Q\mathcal{C}$. Consider now a point $V(1,v_1,v_2,v_3)$ not lying in H_∞. Then V belongs to $P\mathcal{C}$ if and only if

$$\begin{aligned} & (0, a-v_1, b-v_2, c-v_3) \in \mathcal{C} \\ \iff & (a-v_1)(c-v_3)-(b-v_2)^2 = 0 \\ \iff & (ac-b^2) - cv_1 - av_3 + (v_1v_3 - v_2^2) = 0. \end{aligned}$$

A similar computation holds for $V \in Q\mathcal{C}$. Hence $V \in P\mathcal{C} \cap Q\mathcal{C}$ implies that

$$(ac-b^2) - (\alpha\gamma - \beta^2) - (c-\gamma)v_1 - (a-\alpha)v_3 = 0.$$

So V lies in the plane π of equation $((ac-b^2)-(\alpha\gamma-\beta^2))X_0 - (c-\gamma)X_1 - (a-\alpha)X_3 = 0$. Consider the intersection of H_∞ and π. It is the line ℓ satisfying the equations $X_0 = 0$ and $-(c-\gamma)X_1 - (a-\alpha)X_3 = 0$. Clearly, the line ℓ contains the nucleus $N(0,0,1,0)$ and also the point $PQ \cap H_\infty = (0, a-\alpha, b-\beta, c-\gamma)$. $\square$

Remark 5.17. In the previous statement, the intersection points lying in a plane containing N and $PQ \cap H_\infty$ form actually a conic $\mathcal{C}'$. Since the two quadratic cones intersect in an algebraic curve of degree 4 which already contains the conic $\mathcal{C}$, the remaining curve Γ of degree 2 is either a conic $\mathcal{C}'$, either a line with multiplicity 2 or two lines. The last two cases are in contradiction with the fact that P and Q are two distinct points with $PQ \cap H_\infty \notin \mathcal{C} \cup \{N\}$. So it follows that $\Gamma = \mathcal{C}'$.

Corollary 5.18.

Consider two distinct non-adjacent vertices P and Q of $T_2^*(\mathcal{O})$. Their common neighbours are q points lying in a plane containing N and $PQ \cap H_\infty$, and two points on the lines PN and QN.

Proof. Let P and Q be two distinct non-adjacent vertices of $T_2^*(\mathcal{O})$. From the structure of the $\mathrm{GQ}(q-1, q+1)$, P and Q have $q+2$ common neighbours. Consider the lines PN and QN. They intersect only in N. Since P (respectively Q) has a unique projection P' on QN (resp. Q' on PN), P' and Q' are two common neighbours. The q other common neighbours come from the intersection of the two cones $P\mathcal{C}$ and $Q\mathcal{C}$. Hence, from the previous lemma, they lie on a plane containing N and $PQ \cap H_\infty$. □

Theorem 5.19.

The points of three lines of $T_2^*(\mathcal{O})$ containing N and spanning $\mathrm{PG}(3, q)$ form an identifying code of $T_2^*(\mathcal{O})$.

Proof. Consider three lines ℓ_1, ℓ_2, ℓ_3 of $T_2^*(\mathcal{O})$ containing N and spanning $\mathrm{PG}(3, q)$. The points of these lines form a dominating set since any point is either on one of these lines or has a unique projection on each line ℓ_i. As each point has a unique projection on each line ℓ_i, it is clear that two points on these lines are always separated. Similarly, a point incident with a line ℓ_i is always separated from a point not incident with ℓ_1, ℓ_2 or ℓ_3.

Consider now two points S_1 and S_2 that do not lie on the lines ℓ_i. Assume that these points are not separated. In other words, assume that $Q_1 \in \ell_1$, $Q_2 \in \ell_2$ and $Q_3 \in \ell_3$ are common neighbours of S_1 and S_2. If S_1 and S_2 are adjacent, then their common neighbours lie on the same line S_1S_2. Hence Q_1, Q_2 and Q_3 are collinear, a contradiction since ℓ_1, ℓ_2, ℓ_3 span $\mathrm{PG}(3, q)$.

If S_1 and S_2 are not adjacent, then either Q_1, Q_2, Q_3 all lie in the plane containing the nucleus N and $S_1S_2 \cap H_\infty$ (this plane is uniquely defined by the previous corollary), or at least one of them lies in the plane containing S_1, S_2 and N. In the first case, the three points Q_1, Q_2, Q_3 are all in the same plane containing N. Hence, ℓ_1, ℓ_2, ℓ_3 are coplanar which is a contradiction. In the second case, suppose that Q_1 lies in the plane containing S_1, S_2 and N. It follows that Q_1 is incident with the line S_1N or S_2N. It implies that either $S_1 \in \ell_1$ or $S_2 \in \ell_1$, which is a contradiction.

Therefore the set of points on ℓ_1, ℓ_2, ℓ_3 is an identifying code of $T_2^*(\mathcal{O})$. □

Proof of Proposition 5.14. Let ℓ_1, ℓ_2 and ℓ_3 be three lines incident with N and spanning $\mathrm{PG}(3, q)$. Consider the set C consisting of the points of $T_2^*(\mathcal{O})$ on ℓ_1, ℓ_2, ℓ_3. By Theorem 5.19, this set is an identifying code of size $3q$. Let Q_1 be a point on ℓ_1 and Q_2, Q_3 be its projections on respectively ℓ_2 and ℓ_3. The set $C \setminus \{Q_1, Q_2, Q_3\}$ is still a dominating set. Indeed, a point

P that does not lie on the lines ℓ_i can not have Q_1, Q_2 and Q_3 as neighbours. Otherwise, Q_2 would have two projections on the line Q_1P, namely Q_1 and P.

Moreover, we have a one-to-one correspondence between the sets $(N[P] \cap C) \setminus \{Q_1, Q_2, Q_3\}$ and $N[P] \cap C$ since we can easily determine which vertices are eventually missing in the first sets. Hence, $C \setminus \{Q_1, Q_2, Q_3\}$ is an identifying code of $T_2^*(\mathcal{O})$ of size $3q - 3$. □

The next proposition gives lower bounds on the size of an identifying code in any adjacency graph of a $\mathrm{GQ}(q-1, q+1)$. In particular, our previous construction is optimal for $q = 4$ and close to a constant for the other cases.

Proposition 5.20.

Le q be a power of 2. Any identifying code of a $\mathrm{GQ}(q-1, q+1)$ has size at least $3q - 7$. Moreover, it has size at least $8 = 3q - 4$ if $q = 4$, $19 = 3q - 5$ if $q = 8$, $42 = 3q - 6$ if $q = 16$ and $90 = 3q - 6$ if $q = 32$.

Proof. To prove the lower bound on $\gamma^{\mathrm{ID}}(G)$ for an adjacency graph G of a $\mathrm{GQ}(q-1, q+1)$, we use Proposition 4.25. Any identifying code C of a G, with $|C| < q^2 + q - 2$ satisfies the inequality

$$q^3 \leq \frac{|C|^2}{6} + \frac{(2(q^2 + q - 2) + 5)|C|}{6}.$$

Hence, $|C|^2 + (2q^2 + 2q + 1)|C| - 6q^3 \geq 0$. If there exists an identifying code of size $3q - 8$, then the right-hand side of the inequality is equal to

$$(3q - 8)^2 + (2q^2 + 2q + 1)(3q - 8) - 6q^3 = -q^2 - 61q + 56$$

which is negative for all $q \geq 32$. This is a contradiction. Therefore, any identifying code of a $\mathrm{GQ}(q-1, q+1)$ has size at least $3q - 7$. For small values of q, we can obtain a better bound using the same inequality. Since the expression $(3q - c)^2 + (2q^2 + 2q + 1)(3q - c) - 6q^3$ is negative for $(q, c) \in \{(4, 5), (8, 6), (16, 7), (32, 7)\}$, any identifying code of a $\mathrm{GQ}(q-1, q+1)$ has size at least

$$\begin{cases} 8 = 3q - 4 & \text{if } q = 4 \\ 19 = 3q - 5 & \text{if } q = 8 \\ 42 = 3q - 6 & \text{if } q = 16 \\ 90 = 3q - 6 & \text{if } q = 32. \end{cases}$$

□

We can slightly improve the bound for $q = 4$, that is to say for generalized quadrangles $\mathrm{GQ}(3, 5)$. Recall that the adjacency graph of a $\mathrm{GQ}(3, 5)$ has 64 vertices, each vertex belongs to 6 lines of the generalized quadrangle and each line is incident with 4 vertices. The adjacency of a $\mathrm{GQ}(3, 5)$ is depicted in Figure 5.5.

Proposition 5.21.

Any identifying code of a $\mathrm{GQ}(3, 5)$ is of size at least 9.

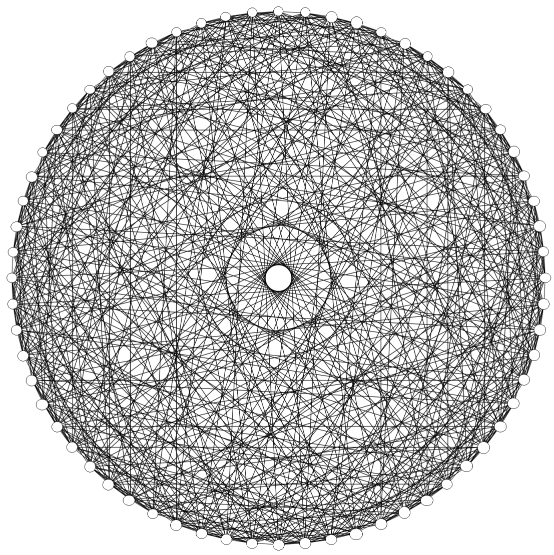


Figure 5.5: Adjacency graph of a GQ(3, 5).

Proof. We use the same discharging method as in Lemma 4.25. Assume that there exists an identifying code C of size 8 of a GQ(3, 5). At the beginning each vertex receives a charge 1. Then each vertex v gives the charge $1/(N[v] \cap C)$ to each vertex of $N[v] \cap C$. The aim is to compute the maximal charge a code vertex can get from its neighbours. Then we prove that the sum of all charges (after discharging) is strictly smaller than 64, which is a contradiction.

Let $c \in C$ be a code vertex. Let n be the number of code vertices in $N(c)$, i.e., $n = |N(c) \cap C|$. In all cases, c gives itself the charge $1/(n+1)$. Consider a line ℓ incident with c and denote by v_1, v_2, v_3 the other vertices of ℓ. Let x be the number of code vertices distinct from c in ℓ. We have $0 \leq x \leq \min(3, n)$ since $x = \#\{v_i \in C \mid i = 1, 2, 3\}$. If $c' \in C$ is a code vertex not adjacent to c, it has a unique projection on ℓ, which is distinct from c. Hence the $7 - n$ vertices of $C \setminus N[c]$ can be partitioned in three parts: one for each vertex v_i of $\ell \setminus \{c\}$. Let n_i denote $|(N(v_i) \setminus \ell) \cap C|$ for $i = 1, 2, 3$. We have

$$n_1 + n_2 + n_3 = 7 - n.$$

The vertex c will receive from the vertices v_i the charge $1/(n_i + 1 + x)$. So c receives from the vertices (distinct from c) of line ℓ a total charge of

$$f(n, x, n_1, n_2) = \frac{1}{n_1 + 1 + x} + \frac{1}{n_2 + 1 + x} + \frac{1}{7 - n - n_1 - n_2 + 1 + x}.$$

We now turn our attention to the possible values of (n, x, n_1, n_2). We have $0 \leq n < 6$, otherwise two of the vertices v_1, v_2, v_3 of a given line ℓ incident with c will not be separated. Moreover, $0 \leq x \leq \min(3, n)$ and $n_3 = 7 - n - n_1 - n_2 \geq 0$. Let $\ell_0, \ldots, \ell_5$ be the six lines incident with c. For each value of n, we consider the partitions $(n_1^{(i)}, n_2^{(i)}, n_3^{(i)})$ of the lines ℓ_i in order to compute the maximal charge of c:

$$\frac{1}{n+1} + \sum_{i=0}^{5} f(n, x^{(i)}, n_1^{(i)}, n_2^{(i)}).$$

If $n = 0$, then $n_i \neq 0$ for any $i = 1, 2, 3$ since $n = 0 = n_i$ implies that c and v_i are not separated. So the possible partitions, up to permutations of the values n_1, n_2, n_3, are given in the following table

(n_1, n_2, n_3)	$(1, 1, 5)$	$(1, 2, 4)$	$(1, 3, 3)$	$(2, 2, 3)$
$f(0, 0, n_1, n_2)$	1.16667	1.03333	1	0.916667

We can use at most three times the partition $(1, 1, 5)$. Indeed, at most seven vertices have c and a unique other code vertex in their neighbourhood. The best choice of the partitions is three lines with partition $(1, 1, 5)$, giving charge $f(0, 0, 1, 1) = 1 + 1/6$ and three lines with partition $(2, 2, 3)$ giving charge $f(0, 0, 2, 2) = 2/3 + 1/4$. In this precise case, c will receive $1 + 3 + 3/6 + 2 + 3/4 = 7.25 < 8$.

If $n = 1$, the possible partitions, up to permutations are given by

(n_1, n_2, n_3)	$(0, 1, 5)$	$(0, 2, 4)$	$(0, 3, 3)$	$(1, 1, 4)$	$(1, 2, 3)$	$(2, 2, 2)$
$f(1, 0, n_1, n_2)$	1.66667	1.53333	1.5	1.2	1.08333	1
$f(1, 1, n_1, n_2)$	0.97619	0.916667	0.9	0.833333	0.783333	0.75

Since $n = 1$, there is exactly one line among $\ell_0, \ldots, \ell_5$ that has one code vertex (i.e., with $x = 1$). This line cannot have a partition $(0, 1, 5)$ (otherwise c and v_1 are not separated). So this line will give charge at most $f(1, 1, 0, 1) = 41/42$. At most one line without code vertex (i.e., $x = 0$) has a partition $(0, 1, 5)$ and gives a charge of at most $f(1, 0, 0, 1) = 5/3$. The other four lines without code vertex give a charge at most $f(1, 0, 1, 1) = 6/5$ that corresponds to a partition $(1, 1, 4)$. Finally, c gets a charge at most $1/2 + 41/42 + 5/3 + 4 \cdot 6/5 = 7.94 < 8$.

If $n = 2$, we have the following partitions and values of $f(2, x, n_1, n_2)$

(n_1, n_2, n_3)	$(0, 1, 4)$	$(0, 2, 3)$	$(1, 1, 3)$	$(1, 2, 3)$
$f(2, 0, n_1, n_2)$	1.7	1.58333	1.25	1.16667
$f(2, 1, n_1, n_2)$	1	0.95	0.866667	0.833333
$f(2, 2, n_1, n_2)$	0.72619	0.7	0.666667	0.65

Either the two code neighbours of c belong to different lines or belong to the same line. In the first case, the best choice of partitions for the two lines with a code vertex, neighbour of c, is $(0, 1, 4)$. So the lines give each a charge at most $f(2, 1, 0, 1) = 1$.

Note that for all lines ℓ_j with $x = 0$, there is at most one vertex $v_i^{(j)}$ with $n_i^{(j)} = 0$ (otherwise some vertices will not be separated). So among the four lines without a code vertex except for c (i.e., with $x = 0$), at most one line has partition $(0, 1, 4)$ and the best choice of partitions for the other three lines is $(1, 1, 3)$. Hence, these four lines give together a charge at most $f(2, 0, 0, 1) + 3f(2, 0, 1, 1) = 17/10 + 3 \cdot 5/4$. So c gets a charge at most $7.78333 < 8$ in this case.

In the second case, the two code neighbours of c belong to the same line. This line cannot have a partition $(0, 1, 4)$ (otherwise, c and v_1 are not separated). So this line will give charge at most $f(2, 2, 1, 1) = 2/3$, corresponding to a partition $(1, 1, 3)$. Then the other five lines can give charge at most $f(2, 0, 0, 1) + 4f(2, 0, 1, 1) = 17/10 + 4 \cdot 5/4$. So c receives a charge at most $7.7 < 8$.

If $n = 3$, the possible partitions and values of of $f(3, x, n_1, n_2)$ are given in the following table

(n_1, n_2, n_3)	$(0, 1, 3)$	$(0, 2, 2)$	$(1, 1, 2)$
$f(3, 0, n_1, n_2)$	1.75	1.66667	1.33333
$f(3, 1, n_1, n_2)$	1.03333	1	0.916667
$f(3, 2, n_1, n_2)$	0.75	0.733333	0.7
$f(3, 2, n_1, n_2)$	0.592857	0.583333	0.566667

If all three code neighbours belong to the same line, say ℓ_0, then the best choice of partitions is $(1,1,2)$ for ℓ_0, once $(0,1,3)$ and three times $(1,1,2)$ for the other lines without a code vertex except of c. So c can get a charge at most

$$1/4 + f(3,3,1,1) + f(3,0,0,1) + 4f(3,0,1,1) = 7.9 < 8.$$

If two code neighbours belong to the same line, say ℓ_0, then the best choice of partitions is $(0,1,3)$ for ℓ_0, $(0,1,3)$ for the line with one code vertex except of c, once $(0,1,3)$ and three times $(1,1,2)$ for the other lines without a code vertex except of c. So c receives a charge at most $1/4 + f(3,2,0,1) + f(3,1,0,1) + f(3,0,0,1) + 3f(3,0,1,1) = 7.78333 < 8$.

Finally, if the three code neighbours belong to three distinct lines, then the best choice of partitions is $(0,1,3)$ for the three lines with a code vertex distinct from c, once $(0,1,3)$ and twice $(1,1,2)$ for the other lines. So c gets a charge at most

$$1/4 + 3f(3,1,0,1) + f(3,0,0,1) + 2f(3,0,1,1) = 7.76667 < 8.$$

It is not possible to have $n = 4$. Indeed, if it is the case, at least two lines, say ℓ_0 and ℓ_1 do not have any code vertex except c. Then at most one vertex on ℓ_0 or ℓ_1 has only c in its neighbourhood. If this vertex is on ℓ_1, it means that ℓ_0 has the partition $(1,1,1)$. But then, ℓ_1 has the partition $(0,1,2)$ and the vertex having only one extra code vertex in its neighbourhood is not separated from one of the vertices of ℓ_0, which is a contradiction.

If $n = 5$, then there are two code vertices in $C \setminus N[c]$. It is impossible to have two lines ℓ_0 and ℓ_1 incident with c that have no other code vertices. Indeed, among the six vertices of ℓ_0 and ℓ_1, two will have the same code vertices in their neighbourhood, which is impossible. So there is exactly one line without another code vertex and five lines with another code vertex. The only partition possible is $(0,1,1)$ up to permutation. So c gets a charge at most $1/6 + f(5,0,0,1) + 5f(5,1,0,1) = 8$.

Note that at least one vertex of C do not have $k = 5$. If c has $k = 5$, then its five neighbours in C are on distinct lines and they all have at most $k = 3$ vertices in their neighbourhood.

In conclusion, after discharging, the sum of all charges is strictly less than 64, the number of vertices of the GQ$(3,5)$. This is a contradiction. □

Corollary 5.22.

For an adjacency graph G of a GQ$(3,5)$, we have

$$\gamma^{\mathrm{ID}}(G) = 9.$$

We thank Nathann Cohen for helping us with implementations in Sage. Thanks to him, we checked by brute force that the optimal size of identifying codes in an adjacency graph of a GQ$(3,5)$ is 9. Moreover, Figure 5.5 depicting an adjacency graph of a GQ$(3,5)$ was produced through these implementations.

5.4.2 Identifying codes in a parabolic quadric which is a GQ(q, q)

Proposition 5.23.

Let q be a prime power. There exists a GQ(q, q) with an identifying code of size $5q - 2 \in \Theta(n^{1/3})$ where n is the number of vertices.

Before giving the proof, we will consider a particular construction of a GQ(q, q) and give some structural properties.

Let q be a prime power. Let Q be the set of points of PG$(4, q)$ that satisfy the equation $X_0^2 + X_1X_2 + X_3X_4 = 0$ (Q is a parabolic quadric).

Lemma 5.24. [HT91, PT84]

The incidence structure $Q(4, q)$ obtained from the points of Q and lines of Q (i.e., lines of PG$(4, q)$ included in Q) is a generalized quadrangle GQ(q, q). Moreover, the closed neighbourhood of a point A of $Q(4, q)$ is exactly the intersection between a hyperplane π_A (the tangent hyperplane) and Q.

Lemma 5.25.

Let A and B be two non-adjacent points of Q. The common neighbours of A and B are coplanar.

Proof. Let π_A (respectively π_B) be the hyperplane containing all the neighbours of A (resp. B). Since A and B are non-adjacent, π_A and π_B are two distinct hyperplanes (of dimension 3). The common neighbours of A and B are all located in the intersection of π_A and π_B which is a plane. □

Proof of Proposition 5.23. We will construct an identifying code for $Q(4, q)$, which is, by Lemma 5.24, a GQ(q, q). Consider a hyperplane $\pi = \mathrm{PG}(3, q)$ intersecting $Q(4, q)$ in a hyperbolic quadric $Q^+(3, q)$ (for example the hyperplane $X_0 = 0$). The hyperbolic quadric is isomorphic to a grid $K_{q+1} \square K_{q+1}$.

Consider three lines ℓ_0, ℓ_1, ℓ_2 of $Q^+(3, q)$ that are pairwise not intersecting. Consider two distinct points $P_1, P_2 \in \ell_2$ and take lines M_1 and M_2 through P_1 and P_2 respectively, both not contained in the $Q^+(3, q)$ and hence not lying in the 3-space π.

The set of $3(q + 1) + 2q = 5q + 3$ points $\mathcal{S} = \ell_0 \cup \ell_1 \cup \ell_2 \cup M_1 \cup M_2$ is an identifying code. Since it contains a whole line, it is a dominating set. A point A on a line N_1 of $\mathcal{S}$ is clearly separated from all the points that are not on N_1 since it is adjacent to all the points of N_1. The point A is also separated from all the other points of N_1 since they have different projection on any line N_2 of $\mathcal{S}$ not intersecting N_1. Hence all the points of $\mathcal{S}$ are separated from all the other points.

Consider now a point of $Q^+(3, q) \backslash \mathcal{S}$. It has exactly three neighbours on ℓ_0, ℓ_1, ℓ_2 (that are collinear). Two points of $Q^+(3, q) \backslash \mathcal{S}$ with the same projections on ℓ_0, ℓ_1, ℓ_2 are necessarily collinear. Hence they have different neighbours on M_1 (if the projection on ℓ_2 is not P_1) or on M_2 (otherwise). Hence any point of $Q^+(3,q)\backslash\mathcal{S}$ has a unique set of neighbours.

A point A not in S has four or five neighbours in $\ell_0 \cup \ell_1 \cup \ell_2 \cup M_1 \cup M_2$. Since A does not lie in $Q^+(3,q)$, the three points on ℓ_0, ℓ_1 and ℓ_2 are not collinear, hence they span a plane, that is contained in π. The only points of M_1 and M_2 that could be contained in this plane are the intersection of M_1 and M_2 with π which is exactly the points P_1 and P_2. Since P_1 and P_2 are both in ℓ_2 they cannot be both in the neighbourhood of A. Finally, the neighbours of A in $\mathcal{S}$ are not coplanar. Using Lemma 5.25, A is separated from all the other vertices.

To conclude the proof, note that as before we can remove a point on each line of $\mathcal{S}$ and still have an identifying code (remove a point on ℓ_0, which does not have P_1 or P_2 as a neighbour, and remove its 4 distinct projections on the other lines). □

Next proposition gives a lower bound on the size of any identifying code of a GQ(q,q). In particular, the order of our previous construction is optimal. The proof is similar to the proof of Proposition 5.20.

Proposition 5.26.

Let q be a prime power. Any identifying code of a GQ(q,q) has size at least $3q-4$.

5.4.3 Identifying codes in an elliptic quadric which is a GQ(q,q^2)

Proposition 5.27.

Let q be a prime power. There exists a GQ(q,q^2) with an identifying code of size $5q \in \Theta(n^{1/4})$ where n is the number of vertices.

Before giving the proof, we will consider a particular construction of a GQ(q,q^2) and give some structural properties.

Let q be a prime power. Let Q be the set of points of PG$(5,q)$ that satisfy the equation $f(X_0,X_1)+X_2X_3+X_4X_5=0$ where $f(X_0,X_1)=dX_0^2+X_0X_1+X_1^2$, $d\in\mathbb{F}_q$, is an irreducible binary quadratic form over $\mathbb{F}_q$ (Q is an elliptic quadric).

Lemma 5.28. [HT91, PT84]

The incidence structure $Q^-(5,q)$ obtained from the $(q^3+1)(q+1)$ points of Q and the $(q^3+1)(q^2+1)$ lines of Q (i.e., lines of PG$(5,q)$ included in Q) is a generalized quadrangle GQ(q,q^2). Moreover, the closed neighbourhood of a point A of $Q^-(5,q)$ is exactly the intersection between a hyperplane π_A (the tangent hyperplane of A) and Q.

Lemma 5.29.

Let A and B be two non-adjacent points of Q. The common neighbours of A and B lie in a 3-dimensional space.

Proof. Let π_A (respectively π_B) be the hyperplane containing all the neighbours of A (resp. B). Since A and B are non-adjacent, π_A and π_B are two distinct hyperplanes (of dimension 4). The common neighbours of A and B are all located in the intersection of π_A and π_B which is a 3-dimensional space. □

Proof of Proposition 5.27. We construct an identifying code for $Q^-(5,q)$ which is a generalized quadrangle $\mathrm{GQ}(q,q^2)$. Consider a line ℓ_0 of $Q^-(5,q)$, take two distinct 3-spaces π_1 and π_2 of $\mathrm{PG}(5,q)$ intersecting each other only in ℓ_0 such that $\pi_i \cap Q^-(5,q) = Q^+(3,q)$. Take two lines ℓ_1, ℓ_2 in $\pi_1 \cap Q^-(5,q)$ such that ℓ_0, ℓ_1 and ℓ_2 are pairwise non-intersecting. Using the geometry, one can always consider two lines ℓ_3, ℓ_4 in $\pi_2 \cap Q^-(5,q)$ such that ℓ_0, ℓ_3 and ℓ_4 are pairwise non-intersecting.

We will prove that the set of $5(q+1) = 5q+5$ points of $\mathcal{S} = \{\ell_i\}_{i=0,\ldots,4}$ is an identifying code. Since $\mathcal{S}$ contains a whole line, the set $\mathcal{S}$ is a dominating set.

A point A on a line N_1 of $\mathcal{S}$ is clearly separated from all the points that are not on N_1 since it is adjacent to all the points of N_1. The point A is also separated from all the other points of N_1 since they have different projections on any line N_2 of $\mathcal{S}$ not intersecting N_1. Hence all the points of $\mathcal{S}$ are separated from all the other points.

Any point of $(\pi_1 \cap Q^-(5,q)) \backslash \mathcal{S}$ has exactly three neighbours on ℓ_0, ℓ_1, ℓ_2 (and these neighbours are collinear). Moreover, two points of $(\pi_1 \cap Q^-(5,q)) \backslash \mathcal{S}$ with the same projection on ℓ_0, ℓ_1, ℓ_2 are necessarily collinear. Hence, they have different neighbours on ℓ_3. It follows that all the points of $(\pi_1 \cap Q^-(5,q)) \backslash \mathcal{S}$ are separated from all the other points. Equivalently, also all the points of $(\pi_2 \cap Q^-(5,q)) \backslash \mathcal{S}$ are separated from all the other points.

A point $P \in Q^-(5,q)$ not in $\pi_1 \cup \pi_2$ has five neighbours in $\mathcal{S}$. Since P does not lie in π_1, the three points on ℓ_0, ℓ_1 and ℓ_2 are not collinear, hence they span a plane of π_1, containing one point of ℓ_0. Since P does not lie in π_2, the three points on ℓ_0, ℓ_3 and ℓ_4 are not collinear, hence they span a plane of π_2, containing one point of ℓ_0. Now it is clear that the five neighbours of P span a 4-space. Using Lemma 5.29 it follows that the point P is separated by $\mathcal{S}$ to all other points.

To conclude the proof, note that as before we can remove a point on each line of $\mathcal{S}$ and still have an identifying code (remove a point on ℓ_0 and remove its 4 distinct projections on the lines $\ell_1, \ell_2, \ell_3, \ell_4$). □

The next proposition gives a lower bound on the size of any identifying code of a $\mathrm{GQ}(q,q^2)$. In particular, the order of our previous construction is optimal. The proof is similar to the proof of Proposition 5.20.

Proposition 5.30.

Let q be a prime power. Any identifying code of a $\mathrm{GQ}(q,q^2)$ has size at least $3q+2$.

5.4.4 Identifying codes in a hermitian variety which is a $\mathrm{GQ}(q^2, q)$

Proposition 5.31.

Let q be a prime power. There exists a $\mathrm{GQ}(q^2, q)$ with an identifying code of size $5q^2 - 2 \in \Theta(n^{2/5})$ where n is the number of vertices.

Before giving the proof, we will consider a particular construction of a $\mathrm{GQ}(q^2, q)$ and give some structural properties.

Let q be a prime power. Let H be the set of points of $\mathrm{PG}(3, q^2)$ that satisfy the equation $X_0^{q+1} + X_1^{q+1} + X_2^{q+1} + X_3^{q+1} = 0$ (H is a Hermitian variety).

Lemma 5.32. [HT91, PT84]

The incidence structure $H(3, q^2)$ obtained from the $(q^3+1)(q^2+1)$ points of H and the $(q^3+1)(q+1)$ lines of H (i.e., lines of $\mathrm{PG}(3, q^2)$ included in H) is a generalized quadrangle $\mathrm{GQ}(q^2, q)$. Moreover, the closed neighbourhood of a point A of $H(3, q^2)$ is exactly the intersection between a plane π_A (the tangent hyperplane of A) and H.

It is well known that the dual of $H(3, q^2)$ is $Q^-(5, q)$, see [PT84, 3.2.3].

Lemma 5.33.

Let A and B be two non-adjacent points of H. The common neighbours of A and B lie on a line.

Proof. Let π_A (respectively π_B) be the hyperplane containing all the neighbours of A (resp. B). Since A and B are non-adjacent, π_A and π_B are two distinct planes. The common neighbours of A and B are all located in the intersection of π_A and π_B which is a line. □

Proof of Proposition 5.31. We construct an identifying code for $H(3, q^2)$ which is a generalized quadrangle $\mathrm{GQ}(q^2, q)$.

Consider three disjoint lines ℓ_0, ℓ_1, ℓ_2, two distinct points $P_1, P_2 \in \ell_0$ and two lines ℓ_1' and ℓ_2' containing P_1 and P_2 respectively, and not intersecting ℓ_1 or ℓ_2. The set

$$\mathcal{S} = \ell_0 \cup \ell_1 \cup \ell_2 \cup \ell_1' \cup \ell_2'$$

of $|\mathcal{S}| = 5q^2 + 3$ points will be an identifying code. Since $\mathcal{S}$ contains a whole line, the set $\mathcal{S}$ is a dominating set.

A point A on a line N_1 of $\mathcal{S}$ is clearly separated from all the points that are not on N_1 since it is adjacent to all the points of N_1. The point A is also separated from all the other points of N_1 since they have different projections on any line N_2 of $\mathcal{S}$ not intersecting N_1. Hence all the points of $\mathcal{S}$ are separated from all the other points.

If two points R and Q have the same neighbourhood on $\{\ell_0, \ell_1, \ell_2\}$, then this neighbourhood consists of collinear points by Lemma 5.33. If the line containing these points also contains P_1, then the projections of R and Q on the line ℓ_2' are different. If the line would contain P_2, then the projections of R and Q on the line ℓ_1' are different. Hence, $\mathcal{S}$ is a separating set.

To conclude the proof, note that as before we can remove a point on each line of $\mathcal{S}$ and still have an identifying code (remove a point on ℓ_1, that is not a neighbour of P_1 nor of P_2, and remove its 4 distinct projections on the lines $\ell_0, \ell_2, \ell_1', \ell_2'$). □

The next proposition gives a lower bound on the size of any identifying code of a $\mathrm{GQ}(q^2, q)$. In particular, the order of our previous construction is optimal. The proof is similar to the proof of Proposition 5.20.

Proposition 5.34.

Let q be a prime power. Any identifying code of a $\mathrm{GQ}(q^2, q)$ has size at least $2q^2 - 2$.

5.5 Conclusion and Perspectives

We provide identifying codes for several vertex-transitive families of graphs which have size of the same order as the fractional value. Since the considered graphs have diameter 2, our results can be extended to locating-dominating sets and to metric dimension, providing constructions of optimal order for such sets in new families of strongly regular graphs.

Paley graphs are an example of a family of graphs for which the optimal order for the size of identifying codes is at a logarithmic factor of the fractional value. However, the fractional value is bounded by a constant. It would be interesting to exhibit a family of graphs for which the fractional value is not constant and the integer value has not the same order.

A last natural question arising from this work is to ask whether there exists a graph G (or a family of graphs) where the quotient $\gamma^{\mathrm{ID}}(G)/\gamma_f^{\mathrm{ID}}(G)$ is neither constant nor logarithmic.

Chapter 6

Constant 2-labellings and an application to (r, a, b)-covering codes

We present in this chapter a joint work with my co-advisor Gravier. We introduce the concept of constant 2-labelling of a weighted graph and show how it can be used to obtain perfect weighted coverings. Roughly speaking, a constant 2-labelling of a weighted graph is a 2-colouring of its vertex set which preserves the sum of the weights of black vertices under some automorphisms. We study this problem on four types of weighted cycles. Our results on cycles allow us to determine (r, a, b)-codes in $\mathbb{Z}^2$ whenever $|a - b| > 4$, $r \geq 2$ and we give the precise values of a and b. This is a refinement of Axenovich's theorem proved in 2003.

Contents

Constant 2-labellings are particular 2-colourings of weighted graphs. For every composition of the colouring with an automorphism of a given group, the sum of the weights of the black vertices must be equal to a constant that depends on the colour of a given particular vertex.

The motivation about introducing constant 2-labellings, that are particular 2-colourings, comes from covering problems in graphs. These coverings are coverings with balls of constant radius satisfying special multiplicity condition. Recall from Definition 4.31, that for positive integers r, a, b, an (r, a, b)-*code* of a graph $G = (V, E)$ is a set $S \subseteq V$ of vertices such that every element of S belongs to exactly a balls of radius r with elements of S as centres and every element of $V \setminus S$ belongs to exactly b balls of radius r with elements of S as centres. In the multidimensional grid $\mathbb{Z}^d$, the construction presented by Dorbec et al. (see Theorem 4.33) leads to periodic $(1, a, b)$-codes. In particular, for the 2-dimensional grid, there exist non-periodic $(1, a, b)$-codes but for such values of a and b, there also exist periodic colourings that are $(1, a, b)$-codes [Puz04]. For a larger radius r, Puzynina showed that every (r, a, b)-code of $\mathbb{Z}^2$ is periodic (Theorem 4.35). The notion of constant 2-labellings comes up as a natural translation of the periodicity of (r, a, b)-codes in the infinite grid $\mathbb{Z}^2$.

Using Axenovich's characterization in terms of diagonal colouring of all (r, a, b)-codes in $\mathbb{Z}^2$ with $r \geq 2$ and $|a - b| > 4$ (Theorem 4.36), we show that the existence of (r, a, b)-codes in the infinite grid is linked with the existence of constant 2-labellings in particular cycles. It turns out that studying only four types of weighted cycles is sufficient to characterize all (r, a, b)-codes with $|a - b| > 4$ and to determine explicitly the possible values taken by the constants a and b. Hence, we obtain a refinement of Axenovich's theorem.

This chapter, which is a joint work with my co-advisor Gravier [GV], is organized as follows. The first section is dedicated to the presentation of constant 2-labellings of weighted graphs in a general framework. Then we focus on the constant 2-labellings in four types of weighted cycles. In Section 6.2, we present projection and folding techniques that link constant 2-labellings to (r, a, b)-codes. Hopefully, these techniques can be applied to other problems involving periodic tilings. In Section 6.3, we apply the projection and folding method to obtain all possible values of constants a and b such that there exist (r, a, b)-codes of $\mathbb{Z}^2$ with $|a - b| > 4$ and $r \geq 2$. Note that to apply this method, the colouring of the grid must satisfy some specific properties. Finally, we suggest directions for future work.

6.1 Constant 2-labellings

Given a graph $G = (V, E)$, a particular vertex $v \in V$, a map $w : V \to \mathbb{R}$ and a subgroup A of the set $Aut(G)$ of all automorphisms of G, a *constant 2-labelling* of G is a mapping $c : V \to \{\mathtt{0}, \mathtt{1}\}$ such that there exist constants a and b satisfying

$$a = \sum_{\{u \in V | c(\xi(u)) = \mathtt{1}\}} w(u), \quad \forall \xi \in A_\bullet \text{ and } b = \sum_{\{u \in V | c(\xi'(u)) = \mathtt{1}\}} w(u), \quad \forall \xi' \in A_\circ.$$

where $A_\bullet = \{\xi \in A \mid c(\xi(v)) = \mathtt{1}\}$ and $A_\circ = \{\xi \in A \mid c(\xi(v)) = \mathtt{0}\}$.

Example 6.1. let $G = (V, E)$ be the graph with $V = \{v_0, \dots, v_4\}$ represented in Figure 6.1. Take $v = v_0, A = Aut(G), w : V \to \mathbb{R}$ and $c : V \to \{\mathtt{0}, \mathtt{1}\}$ defined by $w(v_0) = 3$, $w(v_1) = w(v_3) = 2$, $w(v_2) = w(v_4) = 5$ and $c(v_0) = c(v_3) = c(v_4) = \mathtt{0}, c(v_1) = c(v_2) = \mathtt{1}$. It is clear that c is a constant 2-labelling since A contains only two automorphisms, id and

$$\sigma : v_0 \mapsto v_0;\ v_1 \mapsto v_4;\ v_2 \mapsto v_3;\ v_3 \mapsto v_2;\ v_4 \mapsto v_1.$$

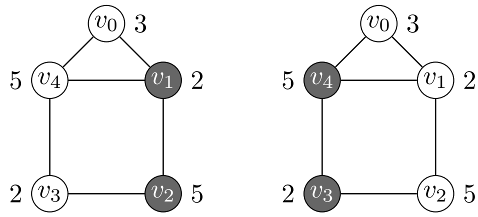


Figure 6.1: A colouring of a graph G and its composition with the automorphism σ.

We can make some straightforward observations about constant 2-labellings. The following proposition allows us to consider either a colouring c or its complement colouring $\overline{c}$.

Proposition 6.2. Complementary property

Let $G = (V, E)$ be a weighted graph, $w : V \to \mathbb{R}$ be the weight map, $v \in V$ and $A \leq Aut(G)$. Set $\omega := \sum_{u \in V} w(u)$. A colouring c is a constant 2-labelling of G with respective constants a and b if and only if the colouring $\overline{c}$ is a constant 2-labelling with respective constants $\omega - b$ and $\omega - a$.

If c is monochromatic black, then the constants are such that $a = \sum_{u \in V} w(u)$ and b is not defined. Otherwise c is white, a is not defined and $b = 0$. It is clear that, for a weighted graph $G = (V, E)$ with $v \in V$, any trivial colouring of V is a constant 2-labelling for any weight map and any subgroup of $Aut(G)$. Such constant 2-labellings are called *trivial*. The definition of constant 2-labellings gives rise to the natural question: whether there exists non-trivial constant 2-labellings for some classes of weighted graphs. We answer that question in the case of four types of weighted cycles in the next subsection.

Remark 6.3. Consider the complete graph K_n and let $w : V(K_n) \to \mathbb{R}$, $v \in V(K_n)$, $A = Aut(K_n)$. It is straightforward to show that there exists a non-trivial constant 2-labelling of K_n if and only if $w(v_1) = w(v_2)$ for all $v_1, v_2 \in V \setminus \{v\}$.

Indeed, assume that $v_1, v_2 \in V \setminus \{v\}$ have different weights and $c : V \to \{\mathtt{0}, \mathtt{1}\}$ is a non-monochromatic colouring. Even if it means taking a composition with an automorphism, we may assume that v_1 and v_2 are of different colours such that $\mathtt{1} = c(v_1) \neq (v_2) = \mathtt{0}$. Let f denote the automorphism that sends v_1 on v_2 and vice versa. We obtain that $c(v) = c \circ f(v)$ and

$$\begin{aligned}
\sum_{c(u)=\mathtt{1}} w(u) &= \sum_{c(u)=\mathtt{1}, u \neq u_1, u_2} w(u) + w(v_1) \\
&\neq \sum_{c(u)=\mathtt{1}, u \neq u_1, u_2} w(u) + w(v_2) = \sum_{(c \circ f)(u)=\mathtt{1}} w(u).
\end{aligned}$$

Therefore, c is not a constant 2-labelling.

Now, suppose that all vertices of $V \setminus \{v\}$ have same weight, say ω. Then any colouring $c : V \to \{\mathtt{0}, \mathtt{1}\}$ of K_n is a constant 2-labelling. Let n be the number of black vertices. It is

clear that
$$\sum_{\xi(u)=1} w(u) = w(v) + (n-1)\omega \quad \text{and} \quad \sum_{\xi'(u)=1} w(u) = n\omega$$
for all $\xi \in A_\bullet, \xi' \in A_\circ$.

6.1.1 Constant 2-labellings in particular weighted cycles

We consider some particular weighted cycles $\mathcal{C}_p$ with at most 4 different weights on the vertices $0, \dots, p-1$. If the weights are $w(0), \dots, w(p-1)$, then we represent the cycle by the word $w(0) \dots w(p-1)$. We will use the letters z, x, y and t to denote the weights of vertices. For instance, the cycle depicted in Figure 6.2 is represented by the word zx^{p-1}.

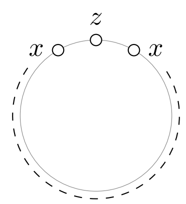


Figure 6.2: Weighted cycle $\mathcal{C}_p$ of Type 0.

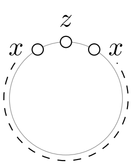
We restrict our study of constant 2-labellings to only four types of weighted cycles (see Figure 6.3) and we set $v := 0$ and $A := \{\mathcal{R}_k \mid k \in \mathbb{Z}\}$. This restriction is due to the initial motivation behind this work: to know the possible values of the constants in (r, a, b)-code of the infinite grid. The four types of weighted cycles on p vertices depend on p mod 4.

- If $p \equiv 1 \pmod 4$, the cycle represented by $z(xy)^{\frac{p-1}{4}}(yx)^{\frac{p-1}{4}}$ is called *Type1mod*.
- If $p \equiv 2 \pmod 4$, the cycle represented by $z(xy)^{\frac{p-2}{4}}t(yx)^{\frac{p-2}{4}}$ is called *Type2mod*.
- If $p \equiv 3 \pmod 4$, the cycle represented by $z(xy)^{\frac{p-3}{4}}xx(yx)^{\frac{p-3}{4}}$ is called *Type3mod*.
- If $p \equiv 0 \pmod 4$, the cycle represented by $z(xy)^{\frac{p-4}{4}}xtx(yx)^{\frac{p-4}{4}}$ is called *Type4mod*.

Hence, we only consider weighted cycles with an axial symmetry in the distribution of weights. It seems to play an important role for the existence of constant 2-labellings. For instance, a weighted cycle $\mathcal{C}_p$ represented by the word $z(xy)^{\frac{p-1}{2}}$ with $x \neq y$, has only monochromatic colourings as constant 2-labellings. See Lemma B.3 given in Appendix B for a proof.

Note that the cycle in Figure 6.2 is a particular case of all of these types. Such cycles are called *Type* 0. As we see in the next lemma, the case of Type 0 cycles is easy to handle.

Lemma 6.4.



For cycles $\mathcal{C}_p$ of Type 0, i.e., zx^{p-1} with $1 < p \in \mathbb{N}$, all colourings are constant 2-labellings.

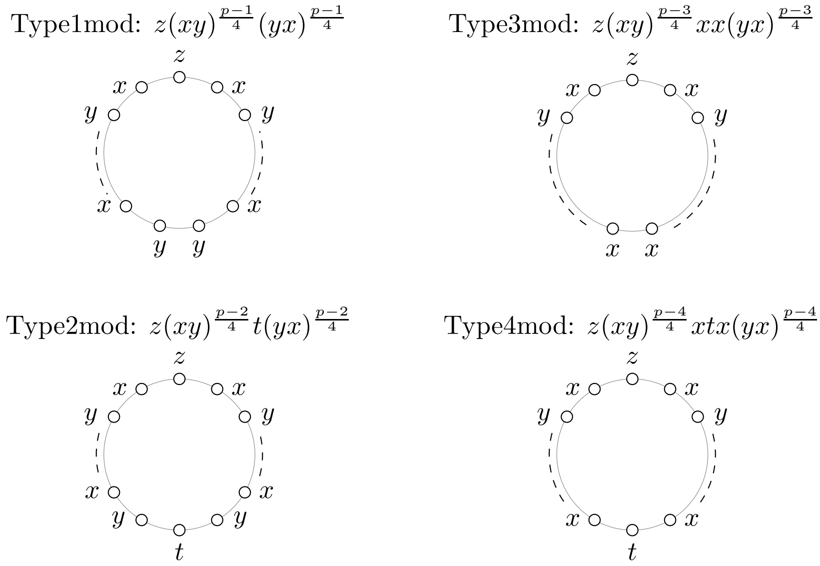
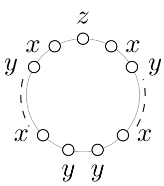


Figure 6.3: Types of weighted cycles $\mathcal{C}_p$.

Proof. Let c be a colouring of $\mathcal{C}_p$. We set α_x to be the number of black vertices with weight x. If $c(0) = \mathtt{1}$, we have $\alpha_x + 1$ black vertices and it is clear that c is a constant 2-labelling where the weighted sum of black vertices is equal to $a = \alpha_x x + z$ (respectively $b = (\alpha_x + 1)x$) if the vertex 0 is black (resp. white). $\square$

We now turn our attention to cycles of Type1mod which are not of Type 0.

Lemma 6.5.



Let $p > 2$ be an integer such that $p \equiv 1 \pmod 4$. For cycles $\mathcal{C}_p$ of Type1mod, i.e., $z(xy)^{\frac{p-1}{4}}(yx)^{\frac{p-1}{4}}$, with $x \neq y$, there exists a non-trivial constant 2-labelling c if and only if $p \equiv 0 \pmod 3$. In which case, c is 3-periodic of pattern period `110`.

Proof. Let $p > 2$ be an integer such that $p \equiv 1 \pmod 4$ and let $\mathcal{C}_p$ be a cycle of Type1mod with $x \neq y$. Assume that c is a non-trivial constant 2-labelling of $\mathcal{C}_p$. The colouring c is not alternate since p is odd. Hence, without loss of generality, we can assume that there exist two consecutive black vertices. Moreover, we can suppose that these vertices are the vertices 0 and 1 of $\mathcal{C}_p$. Indeed, c is a constant 2-labelling if and only if $c \circ \mathcal{R}_j$ is a constant labelling for all $j \in \mathbb{Z}$.

For the colouring c, we let α_x, α_y denote respectively the number of black vertices with weight x and y. We have $a = \alpha_x x + \alpha_y y + z$. We consider the colour of the vertex $\frac{p+1}{2}$.

Assume first that $c(\frac{p+1}{2}) = \mathtt{1}$. Then, for the colouring $c \circ \mathcal{R}_1$, the sum of the weights of black vertices is

$$a = (\alpha_x - 1)y + z + (\alpha_y - 1)x + y = \alpha_y x + \alpha_x y + z$$

since under a 1-rotation, any black vertex with weight x becomes a black vertex of weight y, except for the vertex 1 which becomes the vertex with weight z, and similarly any black vertex with weight y becomes a black vertex of weight x except for the vertex $\frac{p+1}{2}$ which becomes a vertex of weight y. As the weights x and y are distinct, it implies that $\alpha_x = \alpha_y$ (in order to have a sum of black vertices constant and equal to a). We set $\alpha := \alpha_x = \alpha_y$ for a shorter notation.

Let i be the smallest integer in $\{0, \ldots, \frac{p-1}{2} - 1\}$ such that $c(i+1) = \mathtt{0}$ and assume $c(\frac{p+1}{2} + \ell) = \mathtt{1}$ for any $\ell \in \{0, \ldots, i\}$ (otherwise, consider the colouring $c \circ \mathcal{R}_{\frac{p+1}{2}}$ instead of c). Then $c(\frac{p+1}{2} + i + 1) = \mathtt{0}$ as depicted in Figure 6.4. With the colouring $c \circ \mathcal{R}_{i+1}$, we obtain a sum of the weights of black vertices equal to $b = \alpha x + (\alpha+1)y$ (Figure 6.5). To conclude this case, consider the vertex $i+2$ and observe that whatever value is assigned to $c(i+2)$, we obtain a contradiction (Figure 6.6).

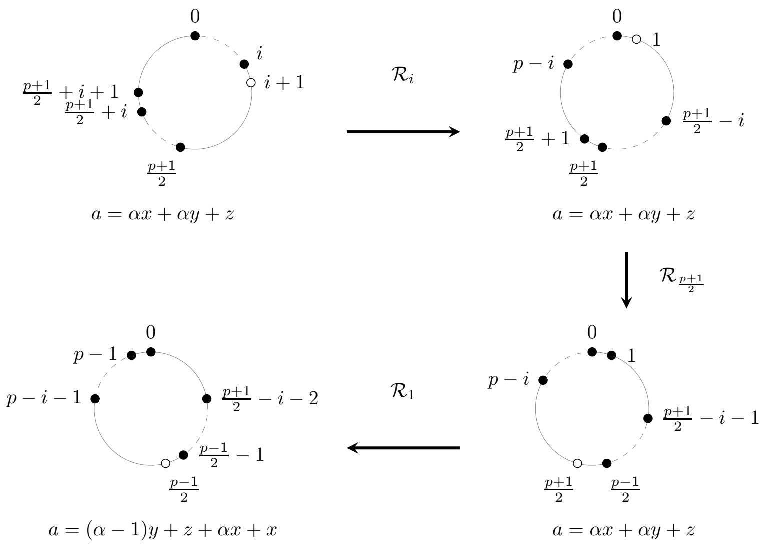


Figure 6.4: Rotations of the colouring c of a Type1mod cycle with $c(\frac{p+1}{2} + i + 1) = \mathtt{1}$, and their corresponding weighted sums of black vertices which are not all equal.

Therefore, we have $c(\frac{p+1}{2}) = \mathtt{0}$ and $a = \alpha_x x + \alpha_y y + z$ as in the beginning. Observe that the previous reasoning means that for any integer j, we have

$$c \circ \mathcal{R}_j(0) = \mathtt{1} = c \circ \mathcal{R}_j(1) \Rightarrow c \circ \mathcal{R}_j\left(\frac{p+1}{2}\right) = \mathtt{0}. \tag{6.1}$$

With the colouring $c \circ \mathcal{R}_1$, the sum of the weights of black vertices is

$$a = (\alpha_x - 1)y + z + \alpha_y x + x = (\alpha_y + 1)x + (\alpha_x - 1)y + z.$$

Since $x \neq y$, we get $\alpha_x = \alpha_y + 1$. We set $\alpha := \alpha_y$ for a shorter notation.

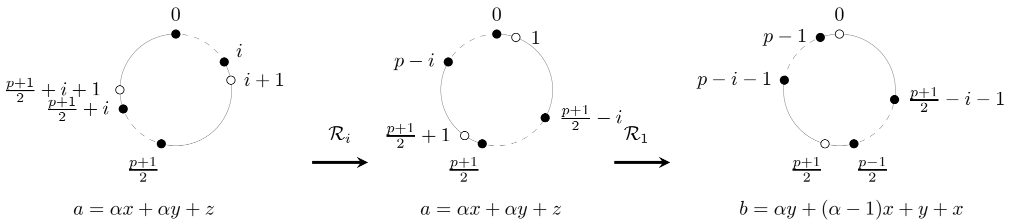


Figure 6.5: Rotations of the colouring c of a Type1mod cycle with $c(\frac{p+1}{2}+i+1)=\texttt{0}$, and their corresponding weighted sums of black vertices.

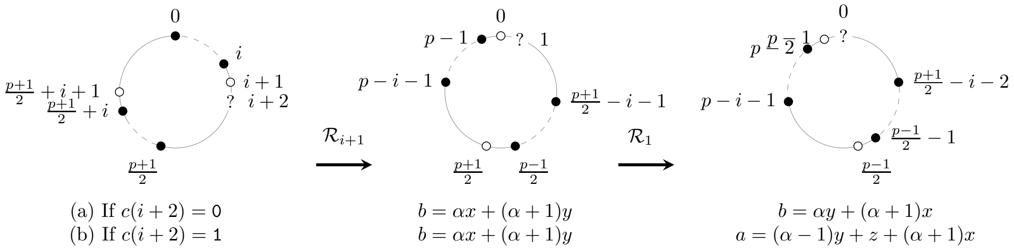


Figure 6.6: Rotations of the colouring c of a Type1mod cycle and their corresponding weighted sums of black vertices depending on the colour $c(i+2)$.

Let i be the smallest integer in $\{0,\ldots,\frac{p-1}{2}-1\}$ such that $c(i+1)=\texttt{0}$. From Equation (6.1), we have $c(\frac{p+1}{2}+\ell)=\texttt{0}$ for any $\ell\in\{0,\ldots,i-1\}$. Moreover, we have $c(\frac{p+1}{2}+i)=\texttt{1}$. Indeed, assume that $c(\frac{p+1}{2}+i)=\texttt{0}$ (Figure 6.7), then with the colouring $c\circ\mathcal{R}_{i+1}$ we obtain a sum of the weights of black vertices equal to $b=(\alpha+1)x+(\alpha+1)y$. As c is a constant 2-labelling, with the colouring $c\circ\mathcal{R}_{\frac{p+1}{2}}$, we have the same weighted sum b. Then it implies that the weighted sum b with the colouring $c\circ\mathcal{R}_{\frac{p+1}{2}+1}$ has a different value, which is a contradiction. So $c(\frac{p+1}{2}+i)=\texttt{1}$ and with the colouring $c\circ\mathcal{R}_{i+1}$, we have a sum of the weights of black vertices equal to $b=\alpha x+(\alpha+2)y$ (Figure 6.8).

From $b=\alpha x+(\alpha+2)y$, it follows that i must be equal to 2, otherwise the colouring $c\circ\mathcal{R}_{\frac{p+1}{2}}$ leads to a different sum of the weights of black vertices (Figure 6.9). Then we have $c(3)=\texttt{1}$ (Figure 6.10). Similarly $c(\frac{p+1}{2}+2)=\texttt{1}$ (Figure 6.11).

Therefore, the colouring $c\circ\mathcal{R}_{\frac{p+1}{2}+1}$ has the same configuration as the colouring c, i.e., the vertices $0,1$ are black and the vertex $\frac{p+1}{2}$ is white. We can apply the same argument as before. Hence, the colouring c must be 3-periodic of pattern period `110` and the number p of vertices is such that $p\equiv 0\pmod 3$. $\square$

Remark 6.6. In the previous proof, we used the following fact. Let c be a constant 2-labelling of a cycle of Type1mod with $x\neq y$. If the sum of the weights of black vertices is equal to $a=\alpha_x x+\alpha_y y+z$ with the colouring c, where α_x, α_y respectively denote the black vertices with weight x and weight y, then for any colouring $c\circ\mathcal{R}_j$ such that $c\circ\mathcal{R}_j(0)=\texttt{1}$, the weighted sum is $a=\alpha_x x+\alpha_y y+z$ and α_x, α_y respectively denote the numbers of black vertices of weight x and weight y with respect to the colouring $c\circ\mathcal{R}_j$. In other words, the number of black vertices of weight x (respectively y) is the same for the colouring $c\circ\mathcal{R}_j$ such

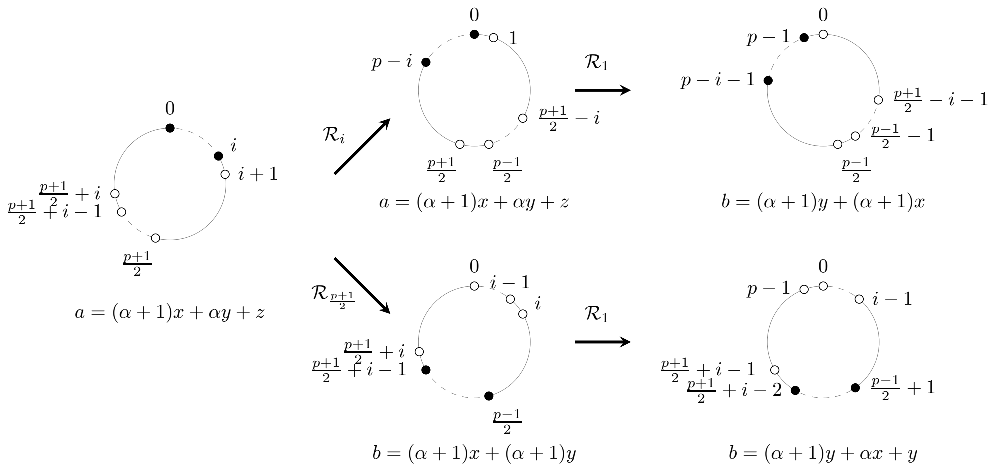


Figure 6.7: Rotations of the colouring c of a Type1mod cycle with $c(\frac{p+1}{2}+i) = \mathtt{0}$, and their corresponding weighted sums b of black vertices which are not all equal.

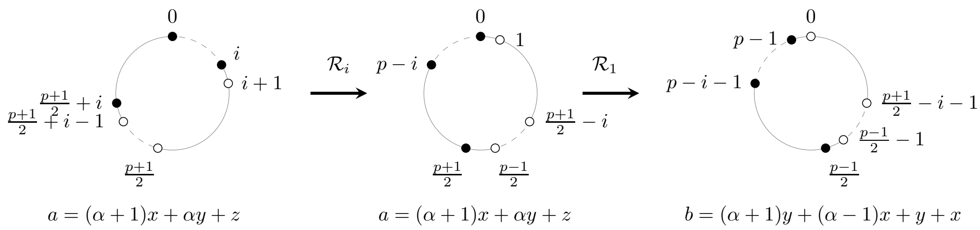


Figure 6.8: Rotations of the colouring c of a Type1mod cycle with $c(\frac{p+1}{2}+i) = \mathtt{1}$, and their corresponding weighted sums of black vertices.

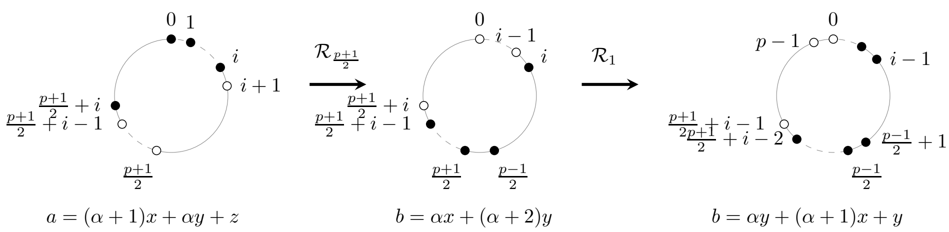


Figure 6.9: Rotations of the colouring c of a Type1mod cycle with $c(j) = \mathtt{1}$ for all $0 \le j \le i$ with $i > 1$, and their corresponding weighted sums of black vertices distinct which are not all equal.

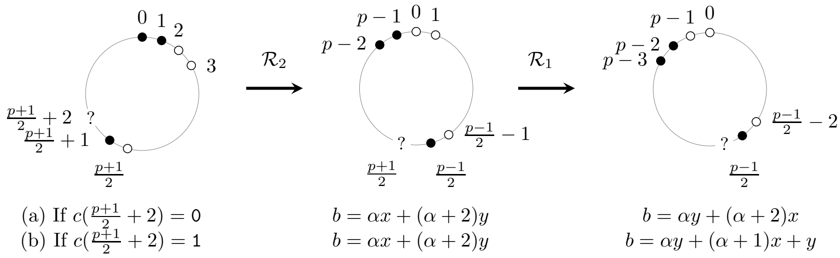


Figure 6.10: Rotations of the colouring c of a Type1mod cycle with $c(0) = c(1) =) = \mathtt{1}$, $c(\frac{p+1}{2}+1) = \mathtt{1}$ and $c(3) = c(\frac{p+1}{2}) = \mathtt{0}$, and their corresponding weighted sums of black vertices depending on the colour $c(\frac{p+1}{2}+2)$.

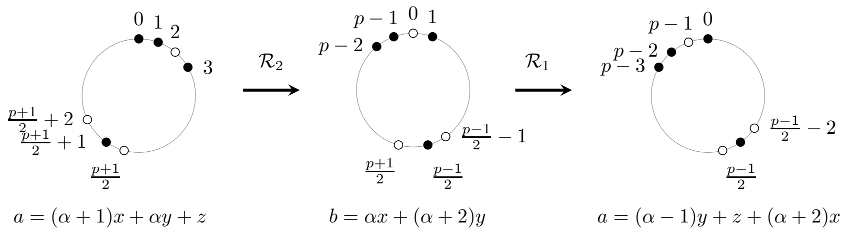


Figure 6.11: Rotations of the colouring c of a Type1mod cycle with $c(0) = c(1) = \mathtt{1}$, $c(\frac{p+1}{2}+1) = \mathtt{1}$ and $c(3) = c(\frac{p+1}{2}) = c(\frac{p+1}{2}+2) = \mathtt{0}$, and their corresponding weighted sums of black vertices which are not all equal.

that $c \circ \mathcal{R}_j(0) = \mathtt{1}$. This fact follows from the uniqueness of the solution (λ, μ) of the system

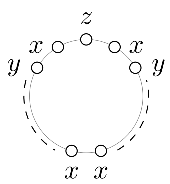
$$\begin{cases} a - z = \lambda x + \mu y \\ n - 1 = \lambda + \mu \end{cases}$$

where n denotes the total number of black vertices (which is known from the colouring c).

The same argument holds when the weighted sum is $b = \alpha_x x + \alpha_y y$ and the colouring $c \circ \mathcal{R}_j$ such that $c \circ \mathcal{R}_j(0) = \mathtt{0}$.

Cycles of Type1mod and Type3mod share some similarities. Both types have at most 3 distinct weights and their non-trivial constant 2-labellings are the same as shown in the next lemma. We omit the proof here since it follows exactly the same lines as the proof of Lemma 6.5, but the details can be found in Appendix B.

Lemma 6.7.



For cycles $\mathcal{C}_p$ of Type3mod, i.e., $z(xy)^{\frac{p-3}{4}}xx(yx)^{\frac{p-3}{4}}$ with $x \neq y$ and $3 < p \in \mathbb{N}$, if c is a non-trivial constant 2-labelling, then $p \equiv 0 \pmod 3$ and c is 3-periodic of pattern period $\mathtt{110}$.

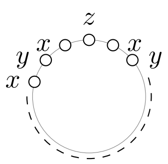
Now for cycles of Type2mod and Type4mod with weights z, x, y, t, if the number n of black vertices and the values $a := z + \alpha_x x + \alpha_y y + \alpha_t t$ and $b := \beta_x x + \beta_y y + \beta_t t$ are known, then the following system

$$\begin{cases} a = \lambda x + \mu y + \nu t + z \\ n = \lambda + \mu + \nu + 1 \end{cases} \qquad \left(\text{respectively} \begin{cases} b = \lambda x + \mu y + \nu t \\ n = \lambda + \mu + \nu \end{cases}\right)$$

does not necessarily have a unique solution $(\lambda, \mu, \nu) = (\alpha_x, \alpha_y, \alpha_t)$ (resp. $(\lambda, \mu, \nu) = (\beta_x, \beta_y, \beta_t)$). Hence for these cycles, it is be important to make a distinction between the colouring c and its rotations.

We first deal with an easy particular case of these cycles that corresponds to a Type2mod cycle with $t = x \neq y$ or to a Type4mod cycle with $t = y \neq x$.

Lemma 6.8.



Let $p \geq 4$ be an integer such that $p \equiv 0 \pmod 2$. Let $\mathcal{C}_p$ be a cycle of of Type2mod with $t = x \neq y$ or of Type4mod with $t = y \neq x$, i.e., $\mathcal{C}_p$ is a cycle represented by $z(xy)^{\frac{p-2}{2}}x$ with $x \neq y$. Any non-trivial constant 2-labelling c of $\mathcal{C}_p$ is either the alternate colouring, or a colouring such that the number α_x of black vertices of weight x is equal to $\alpha_y + c(0)$ where α_y is the number of black vertices of weight y.

Proof. Let $p \geq 4$ be an integer such that $p \equiv 0 \pmod 2$ and let $\mathcal{C}_p$ be a cycle represented by $z(xy)^{\frac{p-2}{2}}x$ with $x \neq y$. Clearly the alternate colouring is a constant 2-labelling with $a = (\frac{p}{2} - 1)y + z$ and $b = \frac{p}{2}x$.

Now assume that c is a non-trivial constant 2-labelling of $\mathcal{C}_p$ which is not the alternate colouring. Without loss of generality, we assume that the vertices 0 and 1 are both coloured in black. Let α_x, α_y denote respectively the number of black vertices with weight x and y for the colouring c. We have $a = \alpha_x x + \alpha_y y + z$ as the sum of the weights of black vertices. For the colouring $c \circ \mathcal{R}_1$, the weighted sum is equal to

$$a = (\alpha_x - 1)y + z + \alpha_y x + x = (\alpha_y + 1)x + (\alpha_x - 1)y + z.$$

As $x \neq y$, we get $\alpha_x = \alpha_y + 1$ and we set $\alpha := \alpha_y$.

Let i be the smallest integer in $\{0, \ldots, p-2\}$ such that $c(i+1) = \mathtt{0}$. The weighted sum for the colouring $c \circ \mathcal{R}_i$ is $a = (\alpha+1)x + \alpha y + z$ by hypothesis. Therefore, the weighted sum for the colouring $c \circ \mathcal{R}_{i+1}$ is equal to $b = (\alpha+1)y + \alpha x + x = (\alpha+1)x + (\alpha+1)y$. Moreover, the weighted sum is preserved for the colouring $c \circ \mathcal{R}_{i+2}$ regardless to the colour of the vertex $i+2$:

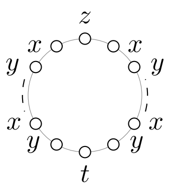
$$\begin{cases} b = (\alpha+1)y + (\alpha+1)x & \text{if } c(i+2) = \mathtt{0} \\ a = (\alpha)y + z + (\alpha+1)y & \text{if } c(i+2) = \mathtt{1}. \end{cases}$$

It follows that the only condition on the constant 2-labelling c is to be a colouring with $\alpha_x = \alpha_y + 1$ if $c(0) = \mathtt{1}$. Similarly, the condition is $\alpha_x = \alpha_y$ if $c(0) = \mathtt{0}$. □

We now consider the Type2mod cycles in general.

Lemma 6.9.

Let $p \equiv 2 \pmod 4$ with $p > 2$ and let $\mathcal{C}_p$ be a weighted cycle of Type2mod represented by $z(xy)^{\frac{p-2}{4}}t(yx)^{\frac{p-2}{4}}$ where the weights x, y, t are not all equal. If c is a non-trivial constant 2-labelling, then c is one of the following colouring



- alternate,
- $\frac{p}{2}$-periodic,
- if $x = y$, $\frac{p}{2}$-anti-periodic,
- if $t = x$, any colouring such that the number of black vertices of weight x is equal to the sum of $c(0)$ and the number of black vertices of weight y.

Proof. Let $p \equiv 2 \pmod 4$ with $p > 2$ and let $\mathcal{C}_p$ be a weighted cycle of Type2mod represented by $z(xy)^{\frac{p-2}{4}}t(yx)^{\frac{p-2}{4}}$ where the weights x, y, t are not all equal. Clearly, the alternate colouring is a constant 2-labelling with $a = (\frac{p}{2} - 1)y + z$ and $b = (\frac{p}{2} - 1)x + t$.

The case where the weights t and x are equal follows from Lemma 6.8. Hence, we suppose from now on that $t \neq x$. Consider a non-trivial constant 2-labelling c of $\mathcal{C}_p$ that is not the alternate colouring. Without loss of generality, we may assume that $c(0) = c(1) = \mathtt{1}$. We let $\alpha_x, \alpha_y, \alpha_t$ denote respectively the number of black vertices of weight x and y for the colouring c. The sum of the weights of the black vertices is then equal to $a = \alpha_x x + \alpha_y y + \alpha_t t + z$. We consider the colour of the vertex $\frac{p}{2} + 1$.

Assume first that $c(\frac{p}{2}+1) = \mathtt{1}$. It follows that $c(\frac{p}{2}) = \mathtt{1}$ and $\alpha_t = 1$, otherwise the weighted sum is not preserved (Figure 6.12). Then for the colouring $c \circ \mathcal{R}_1$, the sum of the weights of the black vertices is

$$a = (\alpha_x - 1)y + z + (\alpha_y - 1)x + t + x + y = \alpha_y x + \alpha_x y + t + z$$

as under a 1-rotation, any black vertex with weight x becomes a black vertex of weight y, except for the vertex 1 which becomes the vertex 0 with weight z, and similarly any black vertex with weight y becomes a black vertex of weight x, except for the vertex $\frac{p}{2}+1$ which becomes the vertex $\frac{p}{2}$ with weight t. If the weights x and y are distinct, then $\alpha_x = \alpha_y$ and we set $\alpha := \alpha_x$. Otherwise, we denote by β the number $\alpha_x + \alpha_y$ of black vertices with weight $x = y$.

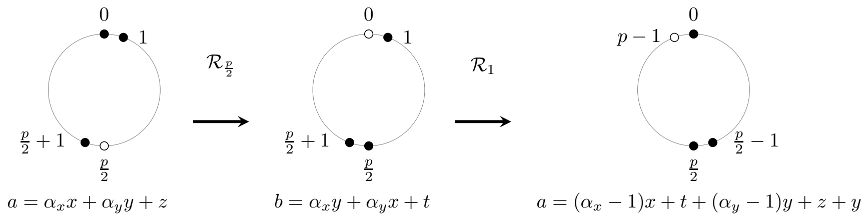


Figure 6.12: Rotations of the colouring c of a Type2mod cycle with $c(\frac{p}{2}+1) = \mathtt{1}$, and their corresponding weighted sums of black vertices which are not all equal as $x \neq t$.

Let i be the smallest integer in $\{0, \ldots, \frac{p}{2}-1\}$ such that $c(i+1) = \mathtt{0}$ and assume that $c(\frac{p}{2}+i) = \mathtt{1}$ for any $\ell \in \{0, \ldots, i\}$ (otherwise, consider the colouring $c \circ \mathcal{R}_{\frac{p}{2}}$ instead of c). Then $c(\frac{p}{2}+i+1) = \mathtt{0}$ as depicted in Figure 6.13.

With the colouring $c \circ \mathcal{R}_{i+1}$ we obtain (Figure 6.14) a sum of the weights of black vertices equal to

$$b = \begin{cases} (\alpha+1)(x+y) & \text{if } x \neq y \\ (\beta+2)x & \text{if } x = y \end{cases}$$

and the number of black vertices of weight x for the colouring $c \circ \mathcal{R}_{i+1}$ is actually $\alpha + 1$ (respectively $\beta + 2$) when $x \neq y$ (resp. $x = y$). Observe that if $c(i+2) = \mathtt{0} = c(\frac{p}{2}+i+2)$, then the weighted sum b for the colouring $c \circ \mathcal{R}_{i+2}$ is preserved.

Therefore, let j be the smallest integer in $\{i+1, \ldots, \frac{p}{2}-1\}$ such that $c(j+1) = \mathtt{1}$. Without loss of generality, we assume that $c(\frac{p}{2}+\ell) = \mathtt{0}$ for all $\ell \in \{i+1, \ldots, j\}$. Then $c\left(\frac{p}{2}+j+1\right) = \mathtt{1}$, otherwise it implies that $x = t$ which is a contradiction (Figure 6.15).

Consequently, the sum of the weights of the black vertices for the colouring $c \circ \mathcal{R}_{j+1}$ is $a = \alpha x + \alpha y + z + t$ (respectively $a = \beta x + t + z$) if the weights x and y are distinct (resp. equal). Moreover, the colourings c and $c \circ \mathcal{R}_{j+1}$ present the same configuration as $c(j+1) = \mathtt{1} = c(\frac{p}{2}+j+1)$ and as the weighted sums are equal. Hence, we can apply the same reasoning given before for c to the colouring $c \circ \mathcal{R}_{j+1}$. It follows that the colouring c is $\frac{p}{2}$-periodic. In particular, we have the following weighted sums

$$\begin{cases} a = \alpha(x+y) + z + t \text{ and } b = (\alpha+1)(x+y) & \text{if } x \neq y \\ a = \beta x + t + z \text{ and } b = (\beta+2)x & \text{if } x = y \end{cases}$$

with β even.

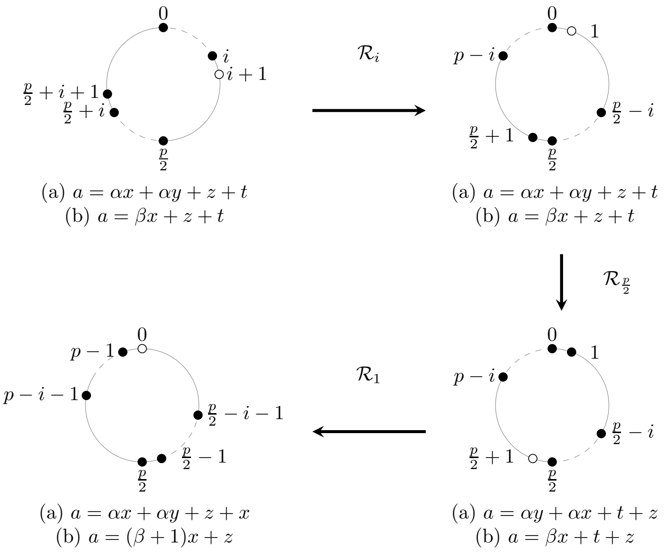


Figure 6.13: Rotations of the colouring c of a Type2mod cycle $\mathcal{C}_p$ with $c(\frac{p}{2}+i+1) = \mathtt{1}$, and their corresponding weighted sums of black vertices which are not all equal, where the line (a) corresponds to the case $x \neq y$ and the line (b) to the case $x = y$.

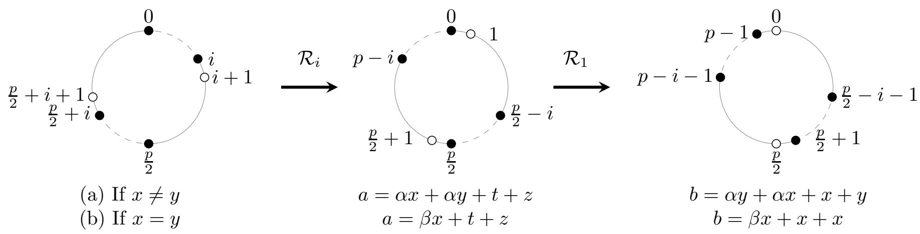


Figure 6.14: Rotations of the colouring c of a Type2mod cycle and their corresponding weighted sums of black vertices depending on the equality of the weights x and y.

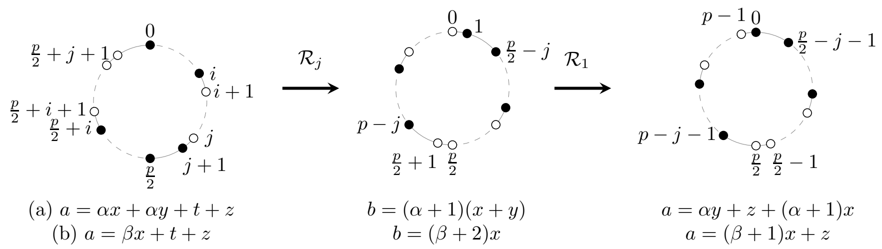


Figure 6.15: Rotations of the colouring c of a Type2mod cycle with $c(j+1) \neq c(\frac{p}{2}+j+1) = \mathtt{0}$, and their corresponding weighted sums of black vertices which are not all equal, where the line (a) corresponds to the case $x \neq y$ and the line (b) to the case $x = y$.

Assume now that $c(\frac{p}{2}+1) = \mathtt{0}$. It follows that $c(\frac{p}{2}) = \mathtt{0}$. Indeed, suppose that $c(\frac{p}{2}) = \mathtt{1}$, i.e., $\alpha_t = 1$. If the weights x and y are distinct, then the weighted sum $a = \alpha_x x + \alpha_y y + t + z$ for the colouring c implies that the weighted sum for the colouring $c \circ \mathcal{R}_{\frac{p}{2}}$ is equal to

$$a = \alpha_x y + \alpha_y x + t + z.$$

Hence, $\alpha_x = \alpha_y$ as c is a constant 2-labelling. Then the weighted sum for the colouring $c \circ \mathcal{R}_1$ is given by

$$a = (\alpha_x - 1)y + z + \alpha_x x + y + x = (\alpha_x + 1)x + \alpha_x y + z$$

which is not equal to the initial weighted sum as $t \neq x$. This is a contradiction. Now, if the weights x and y are equal, then the weighted sum $a = (\alpha_x + \alpha_y)x + t + z$ for the colouring c implies that the weighted sum for the colouring $c \circ \mathcal{R}_1$ is equal to

$$a = (\alpha_x + \alpha_y - 1)x + z + x + x = (\alpha_x + \alpha_y + 1)x + z$$

which is a contradiction (as $t \neq x$).

So $c(\frac{p}{2}+1) = \mathtt{0} = c(\frac{p}{2})$ and $\alpha_t = 0$. For the colouring $c \circ \mathcal{R}_1$, we obtain the weighted sum $a = (\alpha_x - 1)y + z + \alpha_y x + x$ as depicted in Figure 6.16. Hence, α_x must be equal to $\alpha_y + 1$ if the weights x and y are distinct. In this case, we set $\alpha = \alpha_y$. In the case where $x = y$, we simply set $\beta = \alpha_x + \alpha_y$. Hence,

$$\begin{cases} a = (\alpha+1)x + \alpha y + z & \text{if } x \neq y \\ a = \beta x + z & \text{if } x = y. \end{cases}$$

We obtain the following weighted sum (Figure 6.16) for the colouring $c \circ \mathcal{R}_{\frac{p}{2}}$

$$\begin{cases} b = (\alpha+1)y + \alpha x + t & \text{if } x \neq y \\ b = \beta x + t & \text{if } x = y. \end{cases}$$

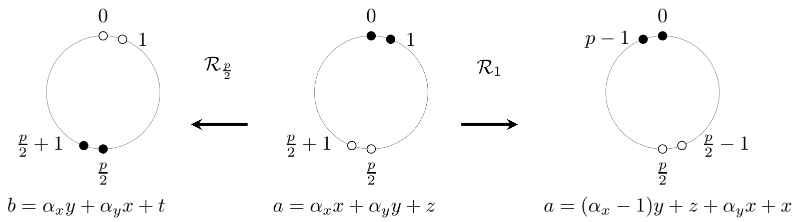


Figure 6.16: Rotations of the colouring c of a Type2mod cycle with $c(\frac{p}{2}) = \mathtt{0} = c(\frac{p}{2}+1)$, and their corresponding weighted sums of black vertices.

Let i be the smallest integer in $\{0, \ldots, \frac{p}{2}-1\}$ such that $c(i+1) = \mathtt{0}$. We may assume that $c\left(\frac{p}{2}+\ell\right) = \mathtt{0}$ for all $\ell \in \{0, \ldots, i\}$. Then $c(\frac{p}{2}+i+1) = \mathtt{1}$, otherwise we obtain a contradiction as $x \neq t$ (Figure 6.17).

Observe that if $c(i+2) = \mathtt{0}$ and $c(\frac{p}{2}+i+2) = \mathtt{1}$, then the weighted sum b for the colouring $c \circ \mathcal{R}_{i+2}$ is preserved. Therefore, let j be the smallest integer in $\{i+1, \ldots, \frac{p}{2}+i\}$ such that $c(j+1) = \mathtt{1}$. Without loss of generality, we suppose that $c\left(\frac{p}{2}+\ell\right) = \mathtt{1}$ for all $\ell \in \{i+1, \ldots, j\}$. It follows that $c\left(\frac{p}{2}+j+1\right) = \mathtt{0}$. Indeed, $c\left(\frac{p}{2}+j+1\right) = \mathtt{1}$ leads to a contradiction as $x \neq t$ (Figure 6.18).

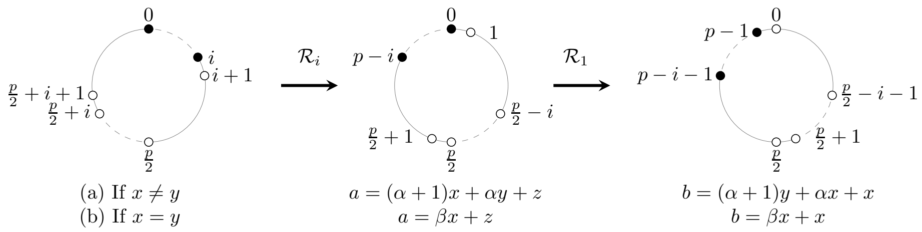


Figure 6.17: Rotations of the colouring c of a Type2mod cycle and their corresponding weighted sums of black vertices depending on the equality of the weights x and y.

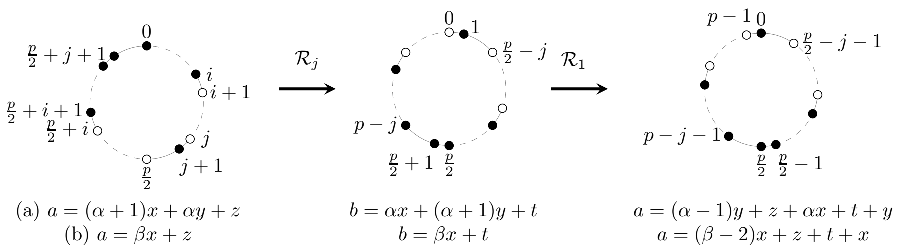


Figure 6.18: Rotations of the colouring c of a Type2mod cycle with $c(j+1) \neq c(\frac{p}{2}+j+1) = \mathtt{0}$, and their corresponding weighted sums of black vertices which are not all equal, where the line (a) corresponds to the case $x \neq y$ and the line (b) to the case $x = y$.

Therefore, the sum of the weights of the black vertices for the colouring $c \circ \mathcal{R}_{j+1}$ is $a = (\alpha+1)x + \alpha y + z$ (respectively $a = \beta x + z$) if the weights x and y are distinct (resp. equal). Hence, the colourings c and $c \circ \mathcal{R}_{j+1}$ present the same configuration as $c(j+1) = \mathtt{1}$ and $c(\frac{p}{2}+j+1) = \mathtt{0}$ and as the weighted sums are equal. It follows that the colouring c is $\frac{p}{2}$-anti-periodic.

If the weights x and y are distinct, then the number of black vertices is equal to $2\alpha+2 = \frac{p}{2}$. It means that $\frac{p}{2}$ is even which is a contradiction as $p \equiv 2 \pmod 4$. Thus, there doest not exist a constant 2-labelling in this case.

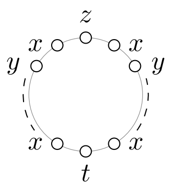
If the weights x and y are equal, then any $\frac{p}{2}$-anti-periodic colouring is a constant 2 labelling with

$$a = \left(\frac{p}{2} - 1\right)x + z \text{ and } b = \left(\frac{p}{2} - 1\right)x + t.$$

□

The last type of cycles is similar to Type2mod cycles. Hence, the proof of the following lemma is similar to the proof of Lemma 6.9. It is thus given in Appendix B.

Lemma 6.10.



Let $p \equiv 4 \pmod 4$ with $p > 4$ and let $\mathcal{C}_p$ be a weighted cycle of Type4mod represented by $z(xy)^{\frac{p-4}{4}} xtx(yx)^{\frac{p-4}{4}}$ where the weights x, y, t are not all equal. If c is a non-trivial constant 2-labelling, then c is one of the following colouring

- alternate,
- $\frac{p}{2}$-anti-periodic,
- $\frac{p}{2}$-periodic if $x = y$; $\frac{p}{2}$-periodic and such that the numbers of black vertices of weight x and y are equal when $c(0) = \mathtt{0}$ if $y \neq x$,
- if $t = \frac{p}{4}x + (1 - \frac{p}{4})y$, c can be moreover such that $c(i) = c(i + \frac{p}{2}) = \mathtt{1}$ for all even $i \in \{0, \ldots, \frac{p}{2} - 1\}$ and $c(i) \neq c(i + \frac{p}{2})$ for all odd $i \in \{0, \ldots, \frac{p}{2} - 1\}$ (up to a 1-rotation).

Using all the previous lemmas, we can now prove our main theorem.

Theorem 6.11.

Let c be a non-trivial constant 2-labelling of a cycle $\mathcal{C}_p$ of Type 0, Type1mod, Type2mod, Type3mod or Type4mod with $A = \{\mathcal{R}_k \mid k \in \mathbb{Z}\}$ and $v = 0$. Let $a = \sum_{\{u \in V \mid c \circ \xi(u) = \mathtt{1}\}} w(u)$ and $b = \sum_{\{u \in V \mid c \circ \xi'(u) = \mathtt{1}\}} w(u)$ for $\xi \in A_\bullet, \xi' \in A_\circ$. Then the possible values of the constants a and b are given in the following table.

Type	Value of a	Value of b	Condition on parameters
0	$\alpha x + z$	$(\alpha + 1)x$	$\alpha \in \{0, \ldots, p - 2\}$
1mod	$\frac{p}{3}x + (\frac{p}{3} - 1)y + z$	$(\frac{p}{3} - 1)x + (\frac{p}{3} + 1)y$	$p \equiv 0 \pmod 3$
3mod	$\frac{p}{3}x + (\frac{p}{3} - 1)y + z$	$(\frac{p}{3} + 1)x + (\frac{p}{3} - 1)y$	$p \equiv 0 \pmod 3$
2mod	$(\frac{p}{2} - 1)y + z$	$(\frac{p}{2} - 1)x + t$	
	$\alpha(x + y) + t + z$	$(\alpha + 1)(x + y)$	$\alpha \in \{0, \ldots, \frac{p}{2} - 1\}$
4mod	$(\frac{p}{2} - 2)y + z + t$	$\frac{p}{2}x$	
	$(2\alpha + 2)x + 2\alpha y + z + t$	$(2\alpha + 2)(x + y)$	$\alpha \in \{0, \ldots, \frac{p}{4} - 1\}$
	$\frac{p}{4}x + (\frac{p}{4} - 1)y + z$	$\frac{p}{4}x + (\frac{p}{4} - 1)y + t$	
	$\frac{p}{2}x + (\frac{p}{4} - 1)y + z$	$\frac{3p}{4}x$	$t = \frac{p}{4}x + (1 - \frac{p}{4})y$
	$(\frac{p}{4} - 1)y + z$	$\frac{p}{4}x$	$t = \frac{p}{4}x + (1 - \frac{p}{4})y$
	$2\alpha x + t + z$	$2(\alpha + 1)x$	$\alpha \in \{0, \ldots, \frac{p}{2} - 2\}$, $x = y$
	$(\alpha + 1)x + \alpha y + z$	$(\alpha + 1)(x + y)$	$\alpha \in \{0, \ldots, \frac{p}{2} - 2\}$, $t = y$

6.2 Projection and folding method

In this section, we present a method that allows us to translate specific colouring problems of the infinite grid in terms of constant 2-labellings of weighted cycles. We first give an example of such problems. Let t and p be integers and let $\mathbf{t} = (t, 1)$, $\mathbf{p} = (p, 0)$. A *frame* is a set of vertices of a given shape (see Figure 6.19) where one of the vertices plays a special role and therefore is called the *center of the frame*. Let a and b be non-negative integers. We consider the problem of deciding whether there exists a 2-colouring c of the infinite grid such that the colouring is periodic with $c(\mathbf{y} + \mathbf{t}) = c(\mathbf{y}) = c(\mathbf{y} + \mathbf{p})$ for any $\mathbf{y} \in \mathbb{Z}^2$, and that each frame contains

- a black vertices if the center of the frame is black,
- b black vertices if the center of the frame is white.

Clearly, if the frames are the balls of radius r, then the problem is the same as determining if there exists an (r, a, b)-covering code of the infinite grid that is periodic of periods $\mathbf{t}$ and $\mathbf{p}$.

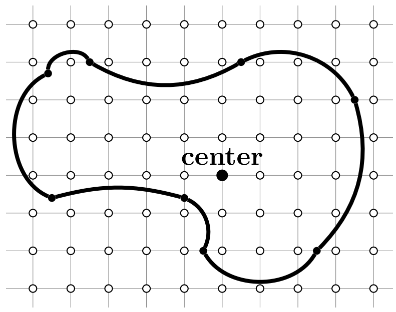


Figure 6.19: An example of a frame shape in the infinite grid.

Now, for $\mathbf{t} = (t, 1)$, $\mathbf{p} = (p, 0)$, consider a 2-colouring of $\mathbb{Z}^2$ that is periodic of periods $\mathbf{t}$ and $\mathbf{p}$. Since c is periodic of period $\mathbf{t}$, the colouring of a line is obtained by doing a translation $\mathbf{t} = (t, 1)$ (respectively -$\mathbf{t} = (-t, -1)$) of the colouring of the line below (resp. above). In this case, if we know the colouring of one line and the translation $\mathbf{t}$, then the colouring of the whole grid $\mathbb{Z}^2$ is known.

Projection

Let $\mathbf{y} \in \mathbb{Z}^2$. Using the translation $\mathbf{t} = (t, 1)$, we can project the frame with center $\mathbf{y}$ on the line L containing $\mathbf{y}$. We assume $\mathbf{y} = (0, 0)$ to simplify the notation. Let $Trans$ denote the set of all the translated frames of the frame with center $\mathbf{y}$ by a multiple of $\mathbf{t}$. Let $h : L \to \mathbb{N}$ be a map defined by

$$h((i, 0)) = \#\{T \in Trans \mid (i, 0) \in T\}.$$

The image of the line L by the mapping h, denoted by $h(L)$, is called the *projection* of the frame with center $\mathbf{y}$ with translation $\mathbf{t} = (t, 1)$. An example is given in Figure 6.20. Observe that $h((i, 0))$ is a finite number for any $i \in \mathbb{N}$, h takes a non-zero value only finitely many times and the number of vertices of a frame is equal to $\sum_{i \in \mathbb{Z}} h((i, 0))$. The map h is introduced to count the number of occurrences in the frame with center $\mathbf{y}$ of vertices of L, up to translation $\mathbf{t}$.

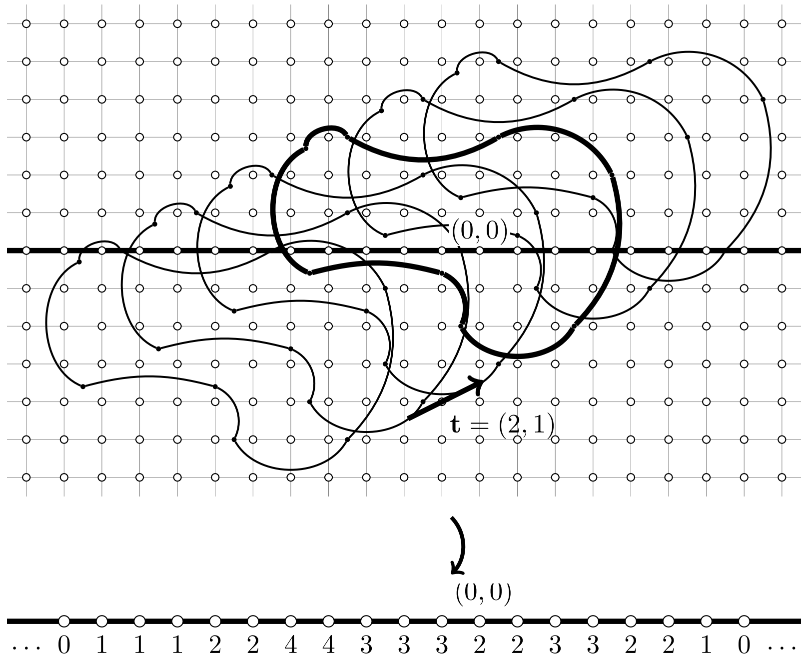


Figure 6.20: Representation of the translated frames with translation $\mathbf{t} = (2, 1)$ of a frame in the finite grid and the projection of this frame on a line.

Folding

Using the translation $(p, 0)$, we can fold a projection on a cycle of p weighted vertices. Let L be the line containing $\mathbf{y} = (0, 0)$ and $\{0, \dots, p-1\}$ be the set of vertices of the cycle $\mathcal{C}_p$. We define a map $w : \{0, \dots, p-1\} \to \mathbb{N}$ such that, for $i \in \{0, \dots, p-1\}$,

$$w(i) := \sum_{k \in \mathbb{Z}} h((i + kp, 0)).$$

The *folding* of the projection $h(L)$ is the cycle $\mathcal{C}_p$ with vertices $0, \dots, p-1$ of respective weights $w(0), \dots, w(p-1)$.

6.3 Application to (r, a, b)-codes of $\mathbb{Z}^2$

The projection and folding method can be used to find (r, a, b)-codes that are periodic. We first give an example, then we characterize the values of a and b of any (r, a, b)-code with $r \geq 2$ and $|a - b| > 4$.

Example 6.12. Consider frames that are balls of radius $r = 3$ and set $t = 2$, $p = 4$. Using the projection and folding method, there exists an (r, a, b)-code of the infinite grid that is periodic of periods $\mathbf{t} = (2, 1)$ and $\mathbf{p} = (4, 0)$ if and only there exists a constant 2-labelling of the cycle $\mathcal{C}_4$ with weights $w(0) = 7, w(1) = w(2) = w(3) = 6$. For instance, the colouring c defined by $c(0) = c(1) = \mathtt{1}$ and $c(2) = c(3) = \mathtt{0}$ is a constant 2-labelling. Hence, there exists an $(3, 13, 12)$-code of $\mathbb{Z}^2$, which is given at the bottom of Figure 6.21.

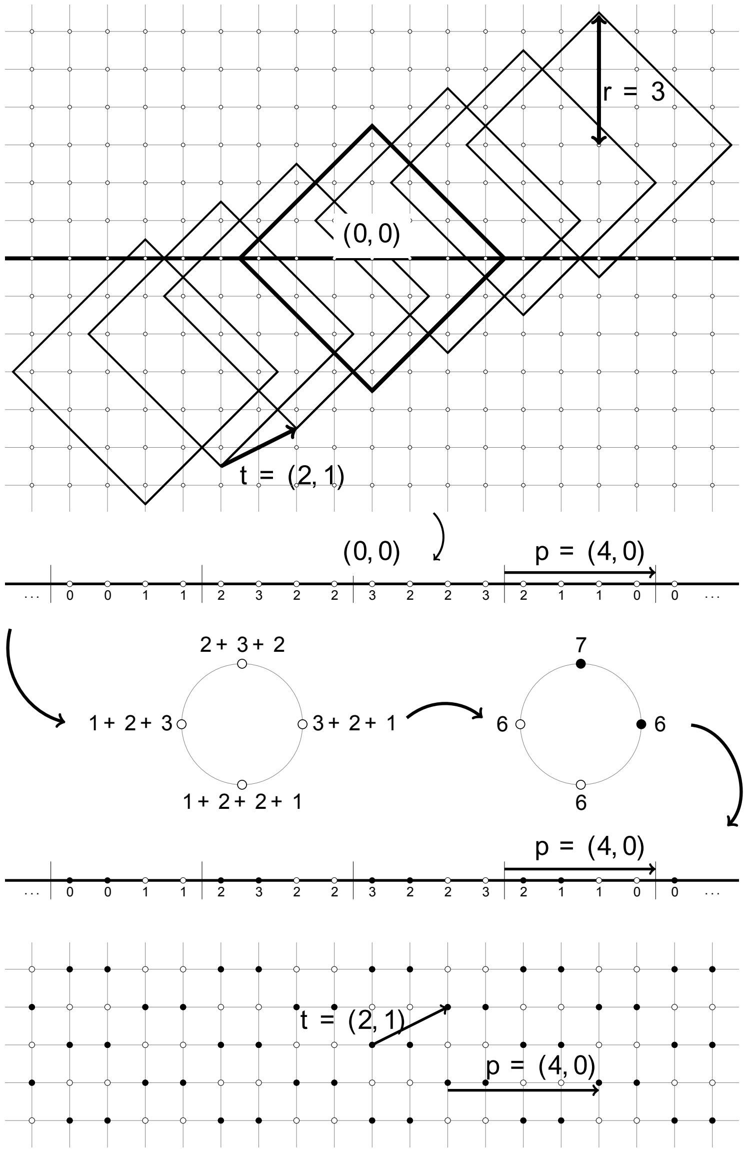


Figure 6.21: Projection and folding of a ball of radius 3 with respect to the translations $\mathbf{t} = (2, 1)$ and $\mathbf{p} = (4, 0)$.

Let $r \geq 2$ and $a, b \in \mathbb{N}$ such that $|a - b| > 4$. Let c be an (r, a, b)-code of $\mathbb{Z}^2$. By Theorem 4.36, c is a diagonal colouring. Hence, c is determined by the colouring of any horizontal line, e.g. $\{(x_1,0) \mid x_1 \in \mathbb{Z}\}$, and by the orientation of the monochromatic diagonals in the even and odd sublattices.

Assume first that the monochromatic diagonals are all parallel. Without loss of generality, we can suppose that they are of the type $\{(x_1, x_1+c) \mid x_1 \in \mathbb{Z}\}$ with $c \in \mathbb{Z}$. Indeed, the case where the monochromatic diagonals are of type $\{(x_1, -x_1+c) \mid x_1 \in \mathbb{Z}\}$ is similar since the grid is symmetric. In this case, if the colouring of a line of $\mathbb{Z}^2$ is known, then the colouring of the line above (resp. below) is obtained by doing a translation $\mathbf{t} = (1,1)$ (resp. $-\mathbf{t}$) as $c(\mathbf{x}) = c(\mathbf{x}+\mathbf{t})$ for all $\mathbf{x} \in \mathbb{Z}^2$. So we can apply the projection method. Moreover, by Theorem 4.35, c is such that $c(\mathbf{x}+(m,0)) = c(\mathbf{x})$ for some $m \in \mathbb{N}$ and all $\mathbf{x} \in \mathbb{Z}^2$. Hence, it is possible to apply the folding method.

Now assume that the monochromatic diagonals are not parallel. We may suppose that the even (resp. odd) sublattice is the union of monochromatic diagonals of type $\{(x_1, x_1+c) \mid x_1 \in \mathbb{Z}\}$ (resp. $\{(x_1, -x_1+c) \mid x_1 \in \mathbb{Z}\}$) with $c \in \mathbb{Z}$. We consider an r-ball $B_r(\mathbf{y})$ with center $\mathbf{y}$. Observe that a diagonal intersecting the ball contains either r or $r+1$ elements of the ball. Moreover two intersecting diagonals belong to the same sublattice. Hence, in terms of counting vertices of a particular colour appearing in the ball, it is equivalent to consider monochromatic diagonals that are parallel or not. So, we can apply the folding method in both cases.

Therefore, for $r \geq 2$ and $|a-b| > 4$, there exists an (r,a,b)-code of the infinite grid $\mathbb{Z}^2$ if and only if there exists a constant 2-labelling of some cycle $\mathcal{C}_p$, with $v = 0$, $A = \{\mathcal{R}_k \mid k \in \mathbb{Z}\}$ and the mapping w defined as before, such that

$$a = \sum_{\{u \in V \mid c\circ\xi(u)=1\}} w(u) \text{ and } b = \sum_{\{u \in V \mid c\circ\xi'(u)=1\}} w(u) \quad \forall \xi \in A_\bullet,\ \xi' \in A_\circ.$$

6.3.1 Characterization of (r,a,b)-codes of $\mathbb{Z}^2$ with $|a-b| > 4$ and $r \geq 2$

Theorem 6.13.

Let $r,a,b \in \mathbb{N}$ such that $|a-b| > 4$ and $r \geq 2$. If there exists an (r,a,b)-code of $\mathbb{Z}^2$, then the values of a and b are given in the following table

a	b	Condition on parameters
$r+1+\alpha(2r+1)$	$(\alpha+1)(2r+1)$	$\alpha \in \{0,\dots,r-1\}$, $r \equiv 0 \pmod 2$
$(r+1)^2 - \alpha(\frac{3r}{2}+1)$	$r^2 + \alpha(\frac{3r}{2}+1)$	$\alpha \in \{0,1\}$, $r \equiv 0 \pmod 2$
$r+1+(\alpha+1)(2r+1)$	$(\alpha+1)(2r+1)$	$\alpha \in \{0,\dots,r-2\}$, $r \equiv 1 \pmod 2$
$r^2 + \alpha\frac{3r+1}{2}$	$(r+1)^2 - \alpha\frac{3r+1}{2}$	$\alpha \in \{0,1\}$, $r \equiv 1 \pmod 2$
$(\alpha+1)\frac{2r^2+2r+2}{3} - 1$	$(\alpha+1)\frac{2r^2+2r+2}{3}$	$\alpha \in \{0,1\}$, $r \equiv 1 \pmod 3$
$(\alpha+1)\frac{2r^2+2r}{3} - \frac{r+1}{3} + 1$	$(\alpha+1)\frac{2r^2+2r}{3} + \frac{r+1}{3}$	$\alpha \in \{0,1\}$, $r \equiv 2 \pmod 3$
$(\alpha+1)\frac{2r^2+2r}{3} + \frac{r}{3} - 1$	$(\alpha+1)\frac{2r^2+2r}{3} - \frac{r}{3}$	$\alpha \in \{0,1\}$, $r \equiv 0 \pmod 3$

Proof. For $r \geq 2$ and $|a-b| > 4$, Axenovich described all possible (r,a,b)-codes (see Theorem 4.36) in terms of diagonal colourings. Theorem 4.36 allows us to apply the projection and folding method in this case. Let $\mathbf{y} = (0, 0)$. We project the ball $B_r(\mathbf{y})$ on the line L using the translation $\mathbf{t} = (1, 1)$ and we obtain for an even (respectively odd) radius r

$$h((i,0)) = \begin{cases} r & \text{if } i \leq r \text{ and } i \text{ is odd} \\ r+1 & \text{if } i \leq r \text{ and } i \text{ is even} \\ 0 & \text{otherwise} \end{cases}$$

and for an odd radius r

$$h((i,0)) = \begin{cases} r+1 & \text{if } i \leq r \text{ and } i \text{ is odd} \\ r & \text{if } i \leq r \text{ and } i \text{ is even} \\ 0 & \text{otherwise.} \end{cases}$$

Indeed, if r is even, then any diagonal of the even (respectively odd) sublattice intersecting the ball contains $r + 1$ (resp. r) elements of $B_r(\mathbf{y})$. The other case can be treated similarly.

Consider now the colourings 1–5 given in Theorem 4.36. For each kind of colouring, we fold the projection of $B_r(\mathbf{y})$ on a cycle $\mathcal{C}_p$, with $p \in \{2, 3, r, r+1, 2r, 2r+1, 2r+2\}$, according to the parity of r (see Table 6.1). Then we use Theorem 6.11 to give the possible values of the constant weighted sums a and b.

The colouring 1 is p-periodic of odd period $p \in \{r, r + 1\}$. Hence it gives two different weighted cycles. If r is even, then $B_r(\mathbf{y})$ is projected and folded on the cycle $\mathcal{C}_{r+1}$ of Type 0 with $z = r + 1$ and $x = 2r + 1$. The corresponding values of the constants are then

$$a = r + 1 + \alpha(2r + 1) \text{ and } b = (\alpha + 1)(2r + 1)$$

with $\alpha \in \{0, \ldots, r - 1\}$. If r is odd, $B_r(\mathbf{y})$ is projected and folded on the cycle $\mathcal{C}_r$ of Type 0 with $z = 3r + 2$ and $x = 2r + 1$. So the corresponding values of the constants are

$$a = 3r + 2 + \alpha(2r + 1) \text{ and } b = (\alpha + 1)(2r + 1)$$

with $\alpha \in \{0, \ldots, r - 2\}$.

The colouring 2 is a p-anti-periodic colouring with $p \in \{r, r+1\}$ and p even. It gives then two different weighted cycles with $2p$ vertices. If r is even, $B_r(\mathbf{y})$ is projected and folded on the cycle $\mathcal{C}_{2r}$ of Type4mod with $z = r+1 = y$, $x = r$ and $t = 2r+2$. Then the corresponding values of the constants are

$$\begin{aligned} a &= \frac{2r}{4}r + (\frac{2r}{4} - 1)(r + 1) + r + 1 = (r + 1)^2 + (\frac{3r}{2} + 1), \\ b &= \frac{2r}{4}r + (\frac{2r}{4} - 1)(r + 1) + 2(r + 1) = r^2 + (\frac{3r}{2} + 1). \end{aligned}$$

If r is odd, $B_r(\mathbf{y})$ is projected and folded on the cycle $\mathcal{C}_{2r+2}$ of Type4mod with $z = r = y$, $x = r + 1$ and $t = 0$. The corresponding values are

$$\begin{aligned} a &= \frac{2r+2}{4}(r + 1) + \frac{2r+2}{4}r = r^2 + \frac{3r+1}{2}, \\ b &= \frac{2r+2}{4}(r + 1) + (\frac{2r+2}{4} - 1)r = (r + 1)^2 + \frac{3r+1}{2}. \end{aligned}$$

The colouring 3 is p-periodic of period $p \in \{r, r + 1\}$ with p even. If r is even, $B_r(\mathbf{y})$ is projected and folded on the cycle $\mathcal{C}_r$. This cycle is a particular case of a Type2mod with $t = x$ or of a Type4mod with $t = y$, according to the value of r mod 4. So $\mathcal{C}_r$ is represented

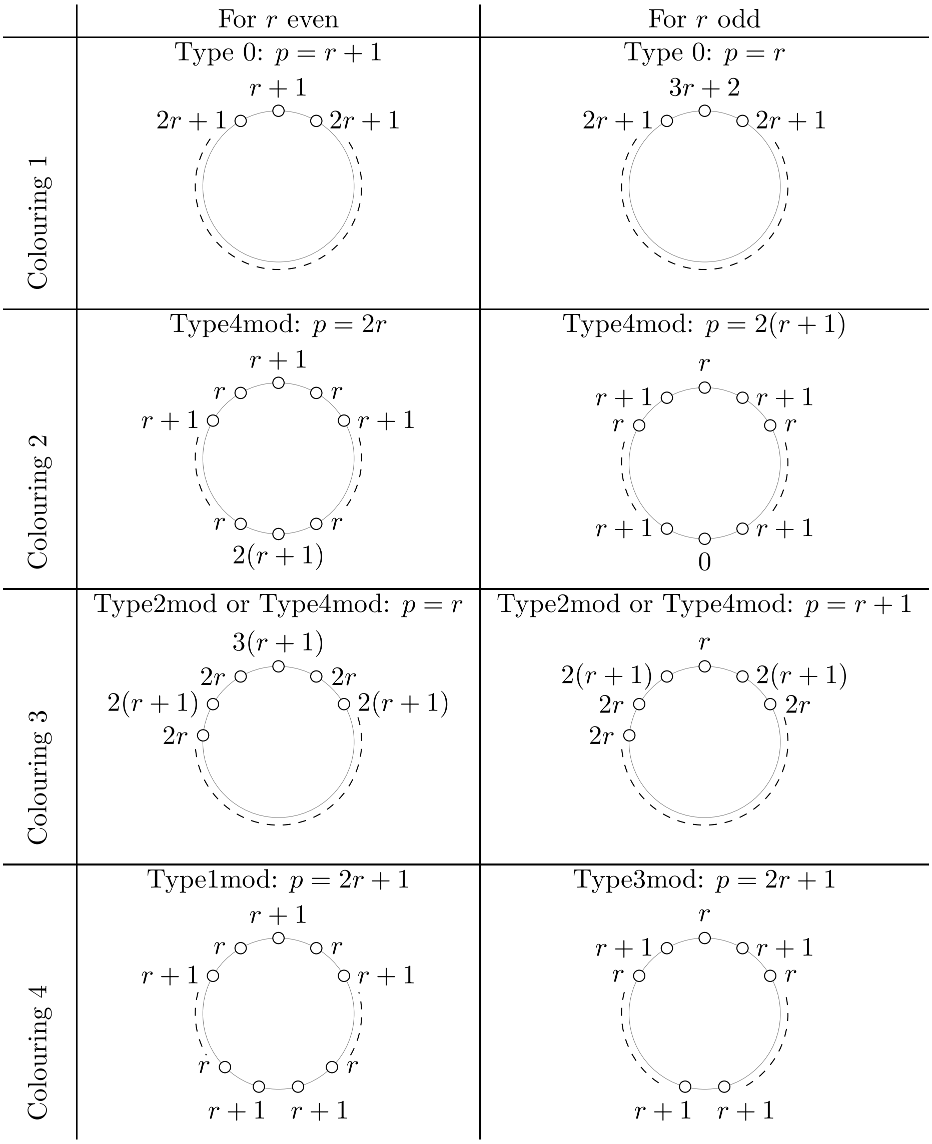


Table 6.1: Weighted cycles $\mathcal{C}_p$ corresponding to the colourings 1–4.

by $z(xy)^{\frac{r-2}{2}}x$ with $z = 3(r+1), x = 2r$ and $y = 2(r+1)$. The corresponding values of the constants are either $a = (r+1)^2$ and $b = r^2$, or

$$a = 2(\alpha+1)r + 2\alpha(r+1) + 3(r+1) \text{ and } b = 2(\alpha+1)(2r+1)$$

with $\alpha \in \{0, \ldots, \frac{r}{2} - 2\}$. Similarly, if r is odd, $B_r(\mathbf{y})$ is projected and folded on the cycle $\mathcal{C}_{r+1}$ which is a particular case of a Type2mod or a Type4mod cycle, represented by $z(xy)^{\frac{r-1}{2}}x$ with $z = r, x = 2(r+1)$ and $y = 2r$. The corresponding values of the constants are either $a = r^2$ and $b = (r+1)^2$ or

$$a = 2(\alpha+1)(r+1) + 2\alpha r + r \text{ and } b = 2(\alpha+1)(2r+1)$$

with $\alpha \in \{0, \ldots, \frac{r-1}{2} - 1\}$.

The colouring 4 is $2r+1$-periodic. If r is even, $B_r(\mathbf{y})$ is projected and folded on the cycle $\mathcal{C}_{2r+1}$ of Type1mod with $z = r+1 = y$ and $x = r$. Such weighted cycle has a constant 2-labelling if $2r+1 \equiv 0 \pmod 3$. Then the corresponding values of the constants are

$$\begin{aligned} a &= \frac{2r+1}{3} \cdot r + \left(\frac{2r+1}{3} - 1\right)(r+1) + r + 1 = \frac{(2r+1)^2}{3} \\ b &= \left(\frac{2r+1}{3} - 1\right) r + \left(\frac{2r+1}{3} + 1\right)(r+1) = \frac{(2r+1)^2}{3} + 1. \end{aligned}$$

If r is odd, $B_r(\mathbf{y})$ is projected and folded on the cycle $\mathcal{C}_{2r+1}$ of Type3mod with $z = r = y$ and $x = r+1$. Hence, under the condition that $2r+1 \equiv 0 \pmod 3$, the corresponding values of the constants are

$$\begin{aligned} a &= \frac{2r+1}{3}(r+1) + \left(\frac{2r+1}{3} - 1\right) r + r + 1 = \frac{(2r+1)^2}{3} + 1 \\ b &= \left(\frac{2r+1}{3} + 1\right)(r+1) + \left(\frac{2r+1}{3} - 1\right) r = \frac{(2r+1)^2}{3} + 1. \end{aligned}$$

Since the difference $|a - b| \leq 4$ in this case, we never obtain these values of a and b.

The colouring 5 is either 2-periodic or 3-periodic. Hence it gives five different weighted cycles. Let c be the colouring 5. If c is 2-periodic, then $B_r(\mathbf{y})$ is projected and folded on $\mathcal{C}_2$ of Type 0, represented by zx with

$$\begin{cases} z = (r+1)^2, x = r^2 & \text{for } r \text{ even} \\ z = r^2, x = (r+1)^2 & \text{for } r \text{ odd.} \end{cases}$$

So the corresponding values of the constants are $a = (r+1)^2$ and $b = r^2$ for r even, and $a = r^2, b = (r+1)^2$ for r odd. If c is 3-periodic, then $B_r(\mathbf{y})$ is projected and folded on $\mathcal{C}_3$ of Type 0. In that case, straightforward analysis give the weights z and x:

- $z = \frac{2r^2+2r-1}{3}$ and $x = \frac{2r^2+2r+2}{3}$ if $r = 3k+1$,
- $z = \frac{2r^2+2r}{3} - 2k + 1$ and $x = \frac{2r^2+2r}{3} + k$ if $r = 3k - 1$,
- $z = \frac{2r^2+2r}{3} + 2k + 1$ and $x = \frac{2r^2+2r}{3} - k$ if $r = 3k$.

The corresponding values of the constants are then given by

$$a = \alpha x + z \text{ and } b = (\alpha+1)x \text{ with } \alpha \in \{0, 1\}.$$

This concludes the proof of Theorem 6.13. $\square$

6.4 Conclusions and perspectives

Constant 2-labellings in weighted cycles allows us to translate the periodicity of (r, a, b)-codes, with $r \geq 2$, of the 2-dimensional grid. It seems that for a radius 1, many $(1, a, b)$-codes of the multidimensional grid $\mathbb{Z}^d$ are periodic (see Theorem 4.33 and [DGHM09]). It would be interesting to find the corresponding weighted graphs obtained with our projection and folding method and then to study constant 2-labellings in these graphs. Also, the projection and folding method is presented in general and can be applied to linear codes. It would be interesting to consider (r, a, b)-codes in other types of lattices as for example, in the king lattice.

The problem of finding a constant 2-labelling of a graph is interesting in and of itself. In Theorem 6.11, we only obtain a characterization of constant 2-labellings in four types of weighted cycles. It would be interesting to consider different weighted cycles, with eventually more weights. Moreover, we could study constant 2-labelling in graphs having a big automorphisms group, for instance, in circulant graphs or in vertex-transitive graphs. Finally, we could find a natural generalization of constant 2-labellings into constant k-labellings using k colours and then consider their links with distinguishing numbers and weighted codes with more than two values.

Appendices

Appendix A

Regularity and ℓ-abelian complexity

We present in this appendix the `Mathematica` code used to compute the 65538 first elements of the 2-abelian complexity of the Thue–Morse word and to conjecture recurrence relations for this complexity. Then we consider the behaviour of abelian and 2-abelian complexity functions of words over a 3-letter alphabet that are generated by 2-uniform morphisms. Finally, we give the proofs omitted in Chapter 2 for the period–doubling word.

Contents

A.1 Mathematica code

In this section, we provide the Mathematica code used to conjecture the relations satisfied by the 2-abelian complexity of the Thue–Morse word. First, we define the four matrices as in Section 2.1 : Me, Mo for factors of length $2n-2$ occurring respectively in even and odd indices, Ne, No for factors of length $2n-1$ occurring respectively in even and odd indices.

The following predicate vectequiv tests whether two given vectors of $\mathbb{N}^{10}$ are similar.

```
:> vectequiv[v_, w_]
      := ((Take[v, 6] - Take[w, 6] == {0, 0, 0, 0, 0, 0})
          && ((v[[7]] + v[[9]] == w[[7]] + w[[9]])
          && (v[[8]] + v[[10]] == w[[8]] + w[[10]])))

:> vectequiv[{0, 1, 0, 1, 1, 1, 0, 0, 0, 1},
             {0, 1, 0, 1, 1, 1, 0, 1, 0, 0}]
Out[]= True
```

The function quotient takes a list of vectors of $\mathbb{N}^{10}$ in argument and outputs the quotient of the given list by the relation $\sim$.

```
:> quotient[ls_]
      := Module[{list = {}, t = ls},
               While[Length[t] > 0,
                     AppendTo[liste, t[[1]]];
                     t = Select[t, ! vectequiv[#, t[[1]]] &];];
               list]

:> quotient[{{0, 1, 0, 1, 1, 1, 0, 0, 0, 1},
             {0, 1, 0, 1, 1, 1, 0, 1, 0, 0},
             {0, 1, 0, 1, 1, 1, 0, 0, 1, 0}}]
Out[]= {{0, 1, 0, 1, 1, 1, 0, 0, 0, 1},
        {0, 1, 0, 1, 1, 1, 0, 0, 1, 0}}
```

We encode the set $S_3 = \{\Psi_2(u) \mid u \in \text{Fac}(\mathbf{t}), |u| = 3\}$ by

```
:> S3 = quotient[{{0, 1, 0, 0, 1, 1, 0, 0, 1, 0},
                  {0, 1, 0, 1, 1, 0, 0, 1, 0, 0},
                  {0, 1, 1, 0, 1, 0, 1, 0, 0, 0},
                  {1, 0, 0, 1, 0, 1, 0, 0, 0, 1},
                  {1, 0, 0, 1, 1, 0, 0, 0, 1, 0},
                  {1, 0, 1, 1, 0, 0, 0, 1, 0, 0}}];
```

Using the tree structure of the sets S_n (Figure 2.4), we create two functions to obtain from a set S_n, its right and left children.

```
:> left[ls_]  := Union[Map[Ne.# &, ls], Map[No.# &, ls]];
:> right[ls_] := Union[Map[Me.# &, ls], Map[Mo.# &, ls]];
```

We are now able to compute the 65538 first values of the 2-abelian complexity $\mathcal{P}_{\mathbf{t}}^{(2)}$ of the Thue–Morse word. These values are given by the list comp, which starts with $1, 2, 4$. We construct comp by iterating two steps: we build a list ls that contains all the sets S_n of a given level in the tree structure using the sets of the previous level, then we append to comp, the cardinal of each quotiented set $S_n/\sim$.

```
:> comp = {1, 2, 4, Length[quotient[S3]]};
:> ls = {S3};
:> For[i = 1, i < 16, i++,
       ls = Flatten[Map[{left[#], right[#]} &, ls], 1];
       comp = Flatten[Append[
                 comp, Map[Length[quotient[#]] &, ls]]];]
```

Remark A.1. This computation of `comp` only takes 4 minutes which is rather fast. Indeed, another approach to compute the 2-abelian complexity of the Thue–Morse word is to generate a long enough prefix of $\mathbf{t}$ and then to count the number of 2-abelian equivalence classes that appear when we slide a window of given length along the prefix. For instance, with the latter method, we must at least iterate the morphism 11 times to find a long enough prefix to compute the first $2^9 + 1 = 513$ first terms. The whole computation already takes more than 4 minutes.

```
:> AbelianFactorCount[list, k, factorlength]
      := Length[Union[
           Table[Sort[Tally[Partition[#, i, 1]]], {i, 1, k}] &
           /@ Partition[list, factorlength, 1]]]
:> AbelianFactorTally[list, k, maxfactorlength]
      := Function[factorlength,
                AbelianFactorCount[list, k, factorlength]]
         /@  Range[0, maxfactorlength]
:> AbsoluteTiming[
      f[x_] := Flatten[x
         /. {"0" -> {"0", "1"}, "1" -> {"1", "0"}}];
      TMword = Nest[f[#] &, {"0"}, 11];
      AbelianFactorTally[TMword, 2, 2^9];]
Out[]= {228.5170705}
```

We consider the 255 first sequences of the 2-kernel of $\mathcal{P}_{\mathbf{t}}^{(2)}$ and encode their prefixes of length 451 in a list called `kernel`. We denote by $\mathbf{x}_{2^e+r}$ the subsequence $\mathcal{P}_{\mathbf{t}}^{(3)}(2^e n + r)_{n\geq 0}$. Then `kernel[[i,j]]` will denote the j^{th} element of the i^{th} sequence of the kernel, $\mathbf{x}_i(j)$.

```
:> kernel
      = Flatten[Table[Table[Table[comp[[2^e n + r + 1]],
              {n, 0, 450}],{r, 0, 2^e - 1}],{e, 0, 7}], 1];
```

Now we try to find a positive integer g such that $\mathbf{x}_1, \dots, \mathbf{x}_g$ are generators of the ideal $\langle \mathcal{K}_2(\mathbf{t})\rangle$. If $2^\ell \leq g + 1 < 2^{\ell+1}$ and $\mathbf{x}_1, \dots, \mathbf{x}_g$ are generators, then any sequence $\mathbf{x}_m$ with $g+1 \leq m \leq 2^{\ell+2}$ must be a linear combination of the generators. For a fixed g, we construct a matrix `M` containing the $g + 1$ first elements of $\mathbf{x}_1, \dots, \mathbf{x}_g$:

$$\mathtt{M} = \begin{pmatrix} \mathbf{x}_1(1) & \mathbf{x}_2(1) & \cdots & \mathbf{x}_g(1) \\ \vdots & \vdots & & \vdots \\ \mathbf{x}_1(g+1) & \mathbf{x}_2(g+1) & \cdots & \mathbf{x}_g(g+1) \end{pmatrix}.$$

```
:> M = Transpose[Table[Table[
         kernel[[i, j]], {j, 1, g + 1}] , {i, 1, g}]];
```

For each $m \in \{g+1, \ldots, 2^{\ell+2}\}$, we create a list `Coefficients` of coefficients c_i (denoted by `c[i]` in the computation) and then we determine the values of coefficients c_i such that

$$\mathbf{x}_m(j) = c_1\mathbf{x}_1(j) + c_2\mathbf{x}_2(j) + \cdots + c_g\mathbf{x}_g(j)$$

for all $j \in \{0, \ldots, g\}$. To obtain a unique solution, we set all free variables to zero using the code `/.c[_]->0`. The solution is then saved in a list called `temp`.

```
:> Coefficients = Table[c[i], {i, 1, g}];
   temp = (
      (Coefficients /. Flatten[Solve
         [(M.Coefficients) == Take[kernel[[m]], g + 1],
          Coefficients]])
       /. c[_] -> 0);
```

Finally, the program prints $\mathbf{x}_m = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_g\mathbf{x}_g$ if the relation guessed from the $g+1$ first elements holds for the 451 first elements. Otherwise, the program prints $\mathbf{x}_m =$ `False`. If at least one `False` appears, then the value of g is too small and has to be changed.

```
:> Print[x[m], " = ",
      If[kernel[[m]] == temp.Table[kernel[[n]], {n, 1, g}],
         temp.Table[x[j], {j, 1, g}], "False"]];
```

Hence, putting everything together, the following program outputs the list of relations given in page 61.

```
:> g = 31;
   M = Transpose[Table[Table[
          kernel[[i, j]], {j, 1, g + 1}] , {i, 1, g}]];
   For[m = g + 1, m < 4 2^Floor[Log[2, g]],
    Quiet[
      Coefficients = Table[c[i], {i, 1, g}];
      temp = (
         (Coefficients /. Flatten[Solve
            [(M.Coefficients) == Take[kernel[[m]], g + 1],
             Coefficients]])
          /. c[_] -> 0);
      Print[x[m], " = ",
       If[kernel[[m]] == temp.Table[kernel[[n]], {n, 1, g}],
        temp.Table[x[j], {j, 1, g}], "False"]];
      m++
      ];
    ]
```

We observe that none of the relations use the sequences $\mathbf{x}_{20}, \ldots, \mathbf{x}_{31}$[1]. Hence, we can reduce the number of generators to $g = 19$. In fact, if we start the "loop `For`" with $m = 1$,

[1]This holds when the program runs under version 7 of `Mathematica`. With the version 9, the program outputs other equivalent relations. For example, $\mathbf{x}_{36} = \mathbf{x}_{21} + \frac{3}{2}\mathbf{x}_{26} - 2\mathbf{x}_{27} + \frac{3}{2}\mathbf{x}_{28} - \frac{3}{2}\mathbf{x}_{29} + \mathbf{x}_{30} - \frac{1}{2}\mathbf{x}_{31}$. In that case, none of the sequences $\mathbf{x}_1, \ldots, \mathbf{x}_{15}$ appear in the relations.

we see from the relations given for $\mathbf{x}_1, \ldots, \mathbf{x}_{63}$ (pages 61 and 61) that only 12 sequences suffice to form a set of generators.

Another possibility to predict the recurrence relations satisfied by a sequence is to use the `Mathematica` package "Regular sequences" created by Rowland in 2010 [Row10]. For a regular sequence, the computation will output either recurrence relations as the one we conjectured, or a set of matrices as in Theorem 1.28.

A.2 Abelian complexity functions satisfying a reflection symmetry

Many abelian complexity functions seem to satisfy a reflection symmetry. First, we consider the abelian complexity of pure morphic words over a 3-letter alphabet $\{\mathtt{0}, \mathtt{1}, \mathtt{2}\}$, that are fixed points of 2-uniform morphisms. Secondly, we consider the 2-abelian complexity of these fixed points.

Without loss of generality, we may assume that the first letter of the fixed points is 0 and then assume that the image of 0 is 01, up to relabelling the letters. Moreover we only consider morphisms that generate an infinite word $\mathbf{w}$ with $|\mathbf{w}|_a \geq 1$ for any $a \in \{\mathtt{0}, \mathtt{1}, \mathtt{2}\}$.

The 1-abelian complexity functions exhibit three distinct behaviours. Some of the functions seem either eventually periodic (Figure A.1), others are not eventually periodic but still satisfy a reflection symmetry in the values taken over each interval $[2^\ell, 2^{\ell+1}]$ for large enough ℓ (Figure A.2). It would be interesting to classify the functions that do not seem to satisfy a reflection symmetry (Figure A.3 and Figure A.4) by their growth rate. For example, the abelian complexity of the word

$$\mathbf{w} = \mathtt{0122111122222222111111111111111122222222222222222222} \cdots$$

generated by the morphism $\mathtt{0} \mapsto \mathtt{01}, \mathtt{1} \mapsto \mathtt{22}, \mathtt{2} \mapsto \mathtt{11}$, seems to grow linearly

$$\mathcal{P}^{(2)}_{\mathbf{w}}(n)_{n \geq \infty} = (1, 3, 4, 5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, \ldots).$$

Consider, for another example, the abelian complexity of the fixed point starting with 0 of the morphism $\mathtt{0} \mapsto \mathtt{01}, \mathtt{1} \mapsto \mathtt{12}, \mathtt{2} \mapsto \mathtt{20}$. The first values of the abelian complexity are

$$1, 3, 6, 7, 12, 12, 13, 12, 18, 19, 21, 18, 19, 18, 21, 19, 27, \ldots$$

and they are depicted in Figure A.4. This complexity satisfies a reflection symmetry over each interval $[2^\ell + 1, 2^{\ell+1} - 1]$ but differs for powers of 2.

In the case of 2-abelian complexity functions, we find the same behaviours again. The functions that seem eventually periodic are depicted in Figure A.5, the ones that seems to satisfy a reflection symmetry are shown in Figure A.6, and the other functions are represented in Figure A.7.

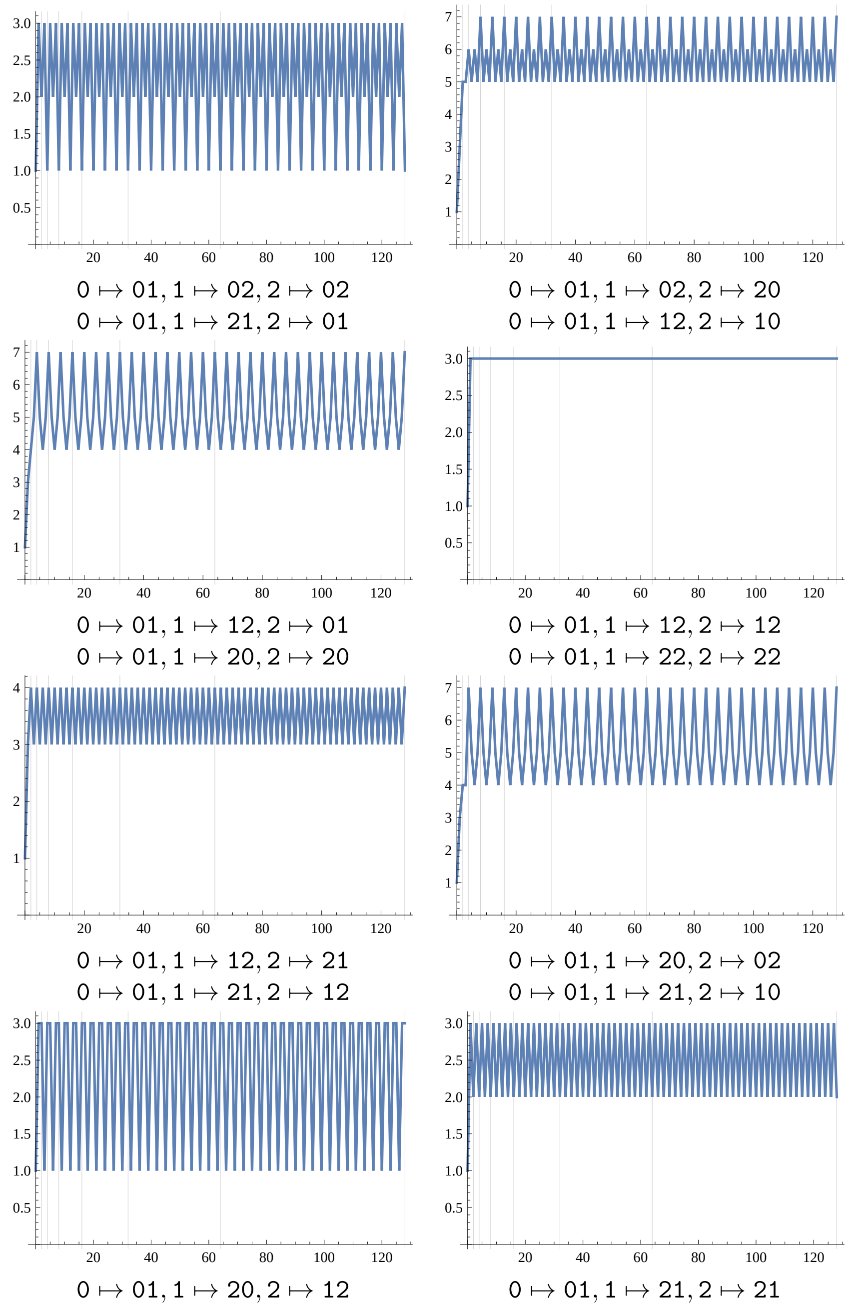


Figure A.1: Abelian complexity functions, which seem eventually periodic, of words generated by a 2-uniform morphism over a 3-letter alphabet

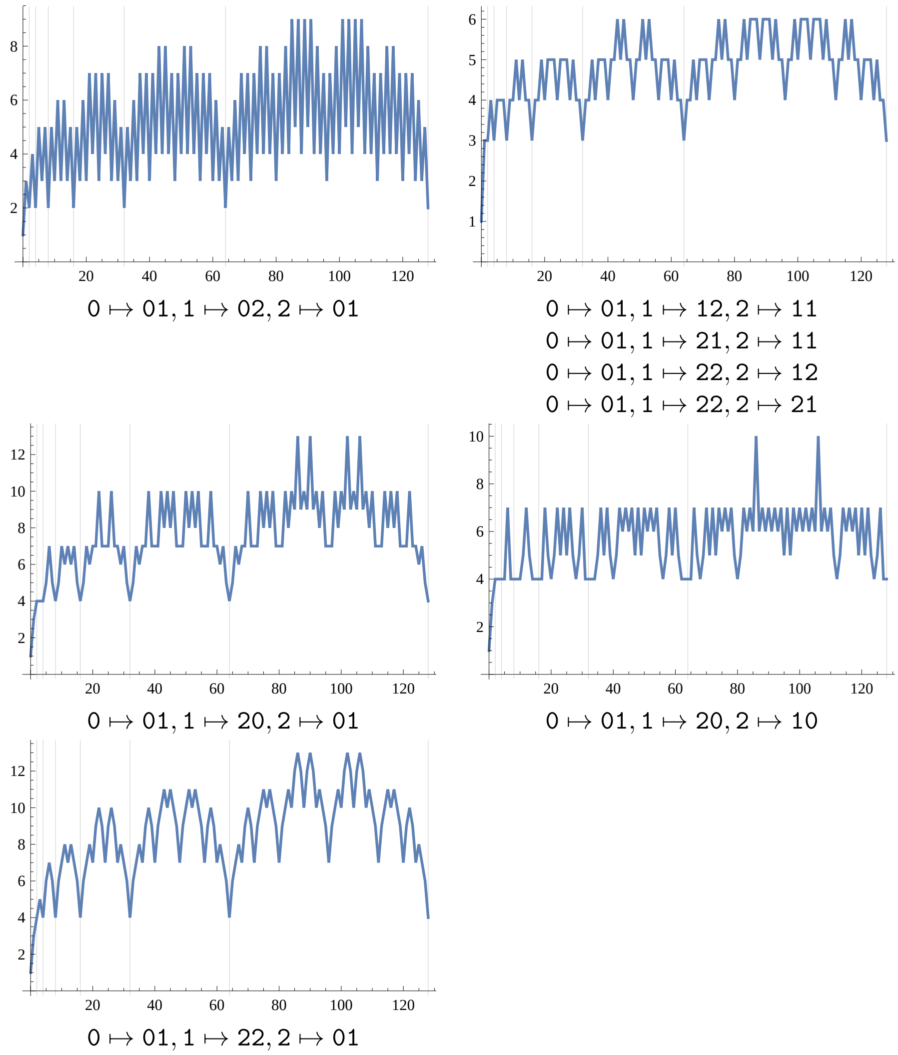
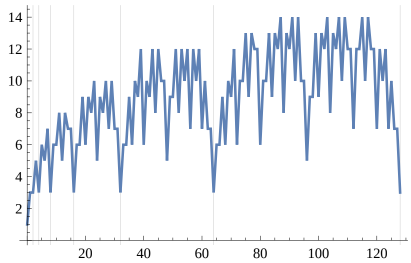
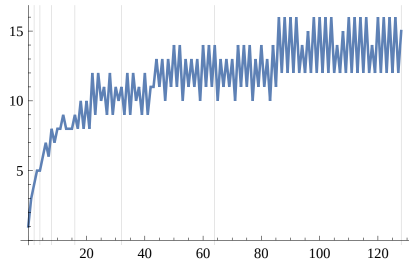


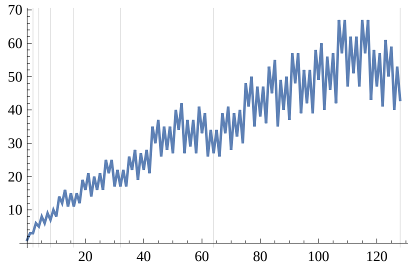
Figure A.2: Abelian complexity functions, which seem to satisfy a reflection symmetry, of words generated by a 2-uniform morphism over a 3-letter alphabet.



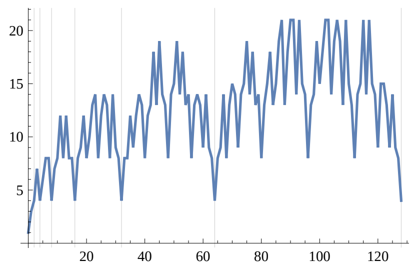
$0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 00$



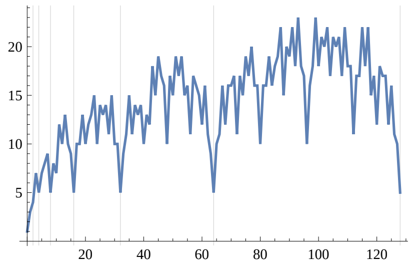
$0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 10$



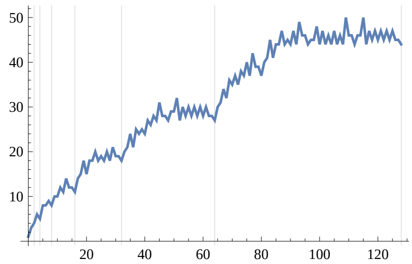
$0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 11$



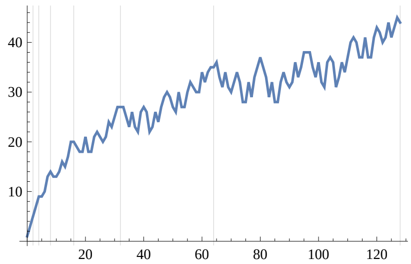
$0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 12$



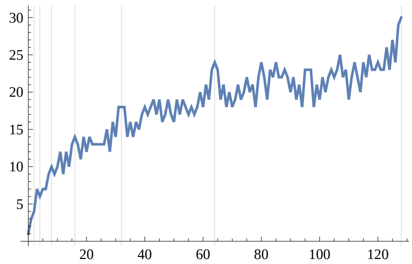
$0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 21$
$0 \mapsto 01, 1 \mapsto 20, 2 \mapsto 21$



$0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 22$



$0 \mapsto 01, 1 \mapsto 12, 2 \mapsto 00$
$0 \mapsto 01, 1 \mapsto 22, 2 \mapsto 20$



$0 \mapsto 01, 1 \mapsto 12, 2 \mapsto 02$
$0 \mapsto 01, 1 \mapsto 21, 2 \mapsto 02$
$0 \mapsto 01, 1 \mapsto 21, 2 \mapsto 20$

Figure A.3: Abelian complexity functions, which do not satisfy a reflection symmetry, of words generated by a 2-uniform morphism over a 3-letter alphabet

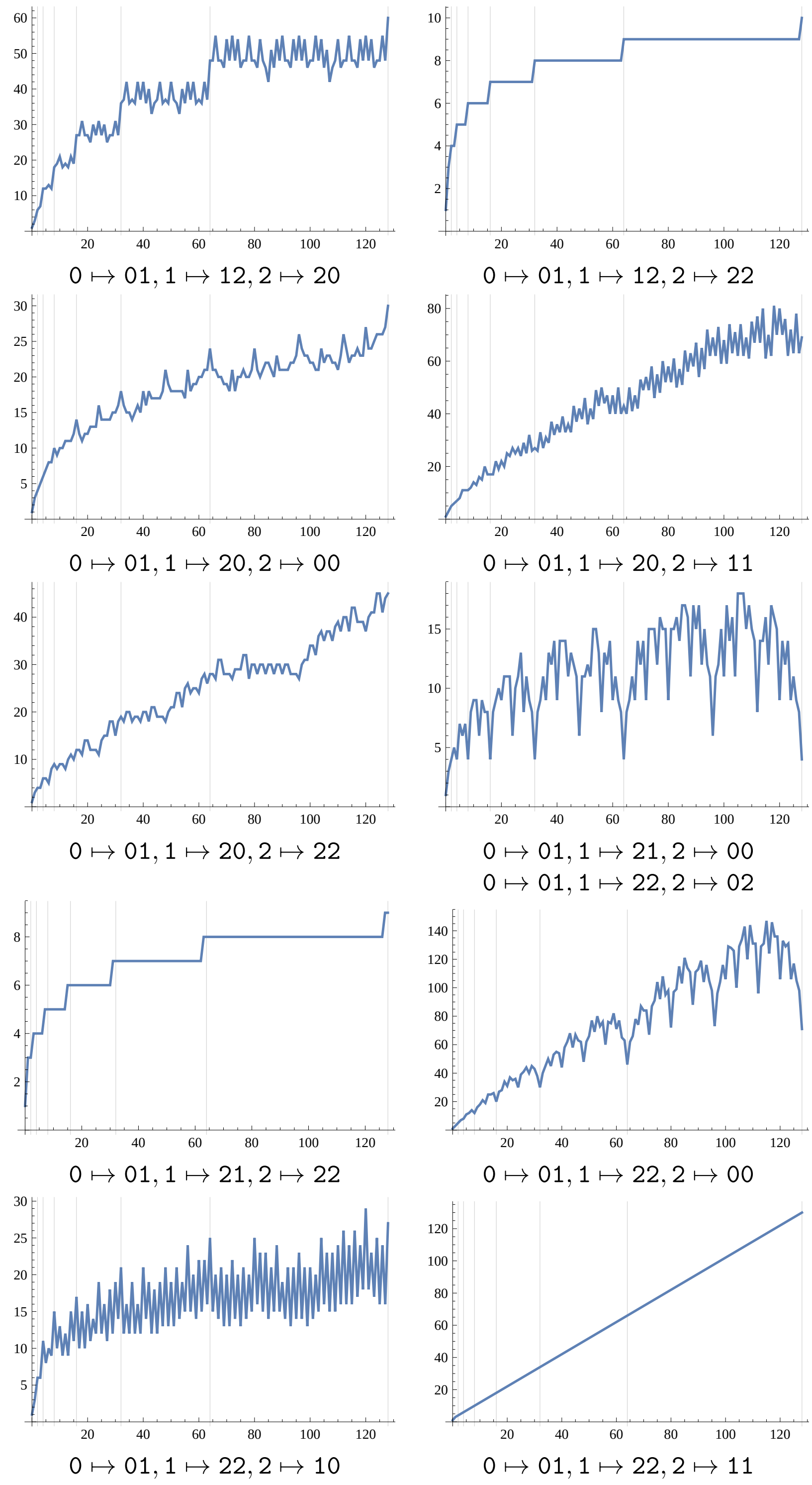


Figure A.4: Abelian complexity functions, which do not satisfy a reflection symmetry, of words generated by a 2-uniform morphism over a 3-letter alphabet

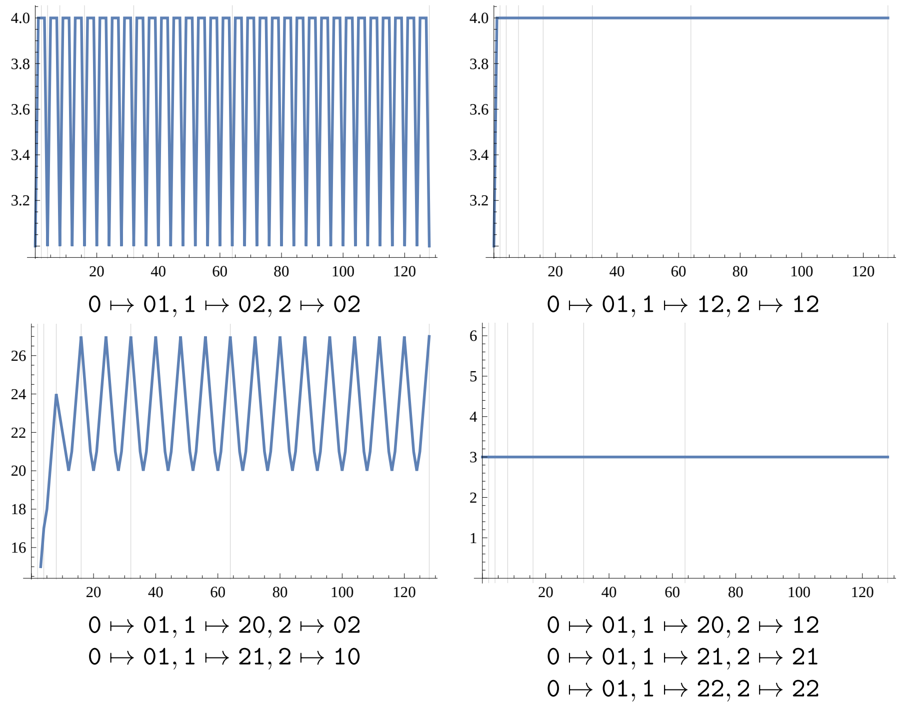


Figure A.5: 2-abelian complexity functions, which seem eventually periodic, of words generated by a 2-uniform morphism over a 3-letter alphabet

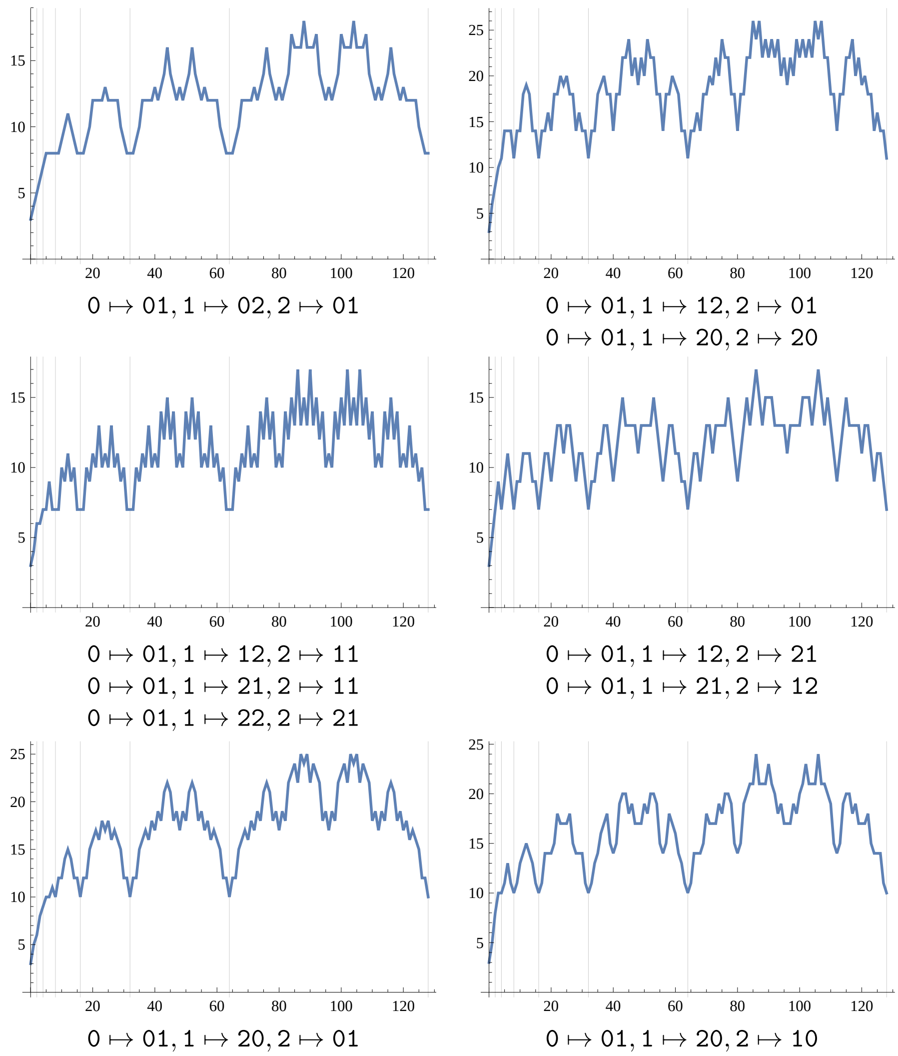
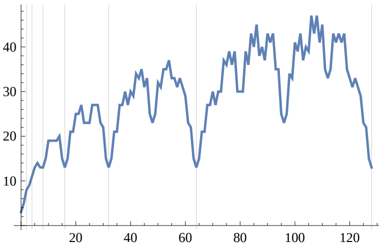
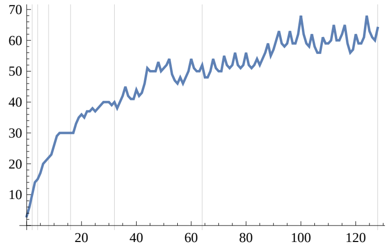


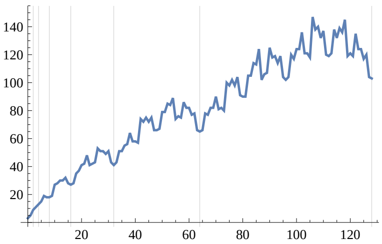
Figure A.6: 2-abelian complexity functions, which seem to satisfy a reflection symmetry, of words generated by a 2-uniform morphism over a 3-letter alphabet



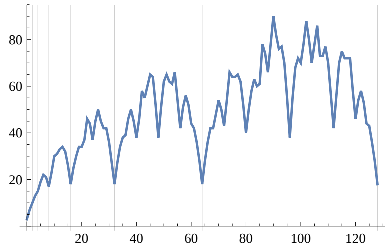
$0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 00$



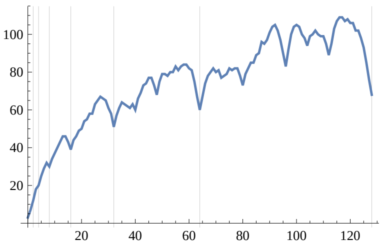
$0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 10$



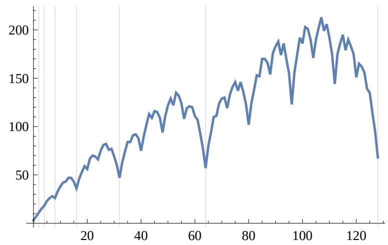
$0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 11$



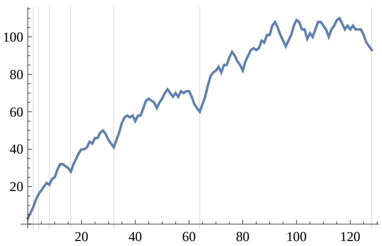
$0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 12$



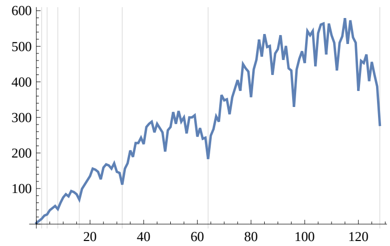
$0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 20$
$0 \mapsto 01, 1 \mapsto 12, 2 \mapsto 10$



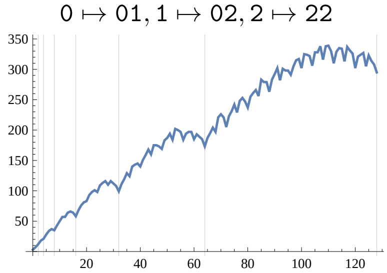
$0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 21$
$0 \mapsto 01, 1 \mapsto 20, 2 \mapsto 21$



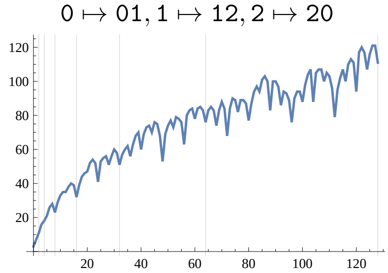
$0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 22$



$0 \mapsto 01, 1 \mapsto 12, 2 \mapsto 20$



$0 \mapsto 01, 1 \mapsto 12, 2 \mapsto 00$
$0 \mapsto 01, 1 \mapsto 22, 2 \mapsto 20$



$0 \mapsto 01, 1 \mapsto 12, 2 \mapsto 02$
$0 \mapsto 01, 1 \mapsto 21, 2 \mapsto 02$
$0 \mapsto 01, 1 \mapsto 21, 2 \mapsto 20$

Figure A.7: 2-abelian complexity functions, which do not satisfy a reflection symmetry, of words generated by a 2-uniform morphism over a 3-letter alphabet

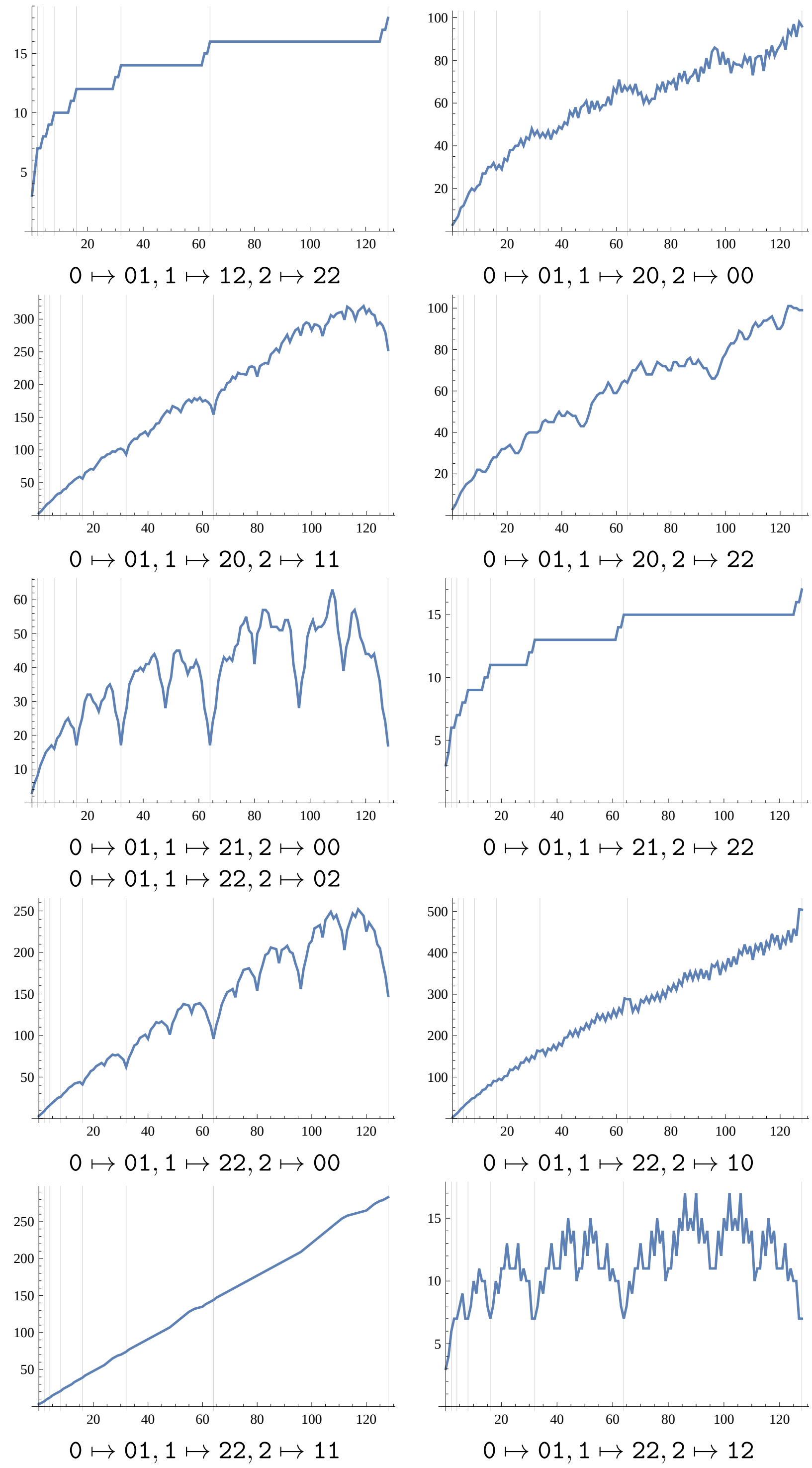


Figure A.8: 2-abelian complexity functions, which do not satisfy a reflection symmetry, of words generated by a 2-uniform morphism over a 3-letter alphabet

A.3 The case of the period-doubling word p

We give in this section the proofs omitted in Chapter 2 for the period–doubling word.

A.3.1 The abelian complexity of $\mathrm{block}(\mathbf{p}, 2)$ is piecewise-defined

In order to prove Proposition 2.39, establishing that $\mathcal{P}_{\mathbf{x}}^{(1)}$ is piecewise-defined in terms of Δ_0, we first mention some properties of factors of the word $\mathbf{x}$.

Lemma A.2.

The set $\mathrm{Fac}_2(\mathbf{x})$ of factors of length 2 occurring in $\mathbf{x}$ is $\{\mathtt{00}, \mathtt{01}, \mathtt{12}, \mathtt{20}, \mathtt{21}\}$.

Proof. It is easy to check that these five words are factors:

$$\mathrm{block}(\mathbf{p}, 2) = \mathtt{12}\underline{\mathtt{00}}\mathtt{1212120}\underline{\mathtt{01}}\mathtt{200}\underline{\mathtt{12}}\mathtt{0012121}\underline{\mathtt{20}}\mathtt{01}\underline{\mathtt{21}}\mathtt{2}\cdots$$

To prove that they are the only ones, it is enough to check that for any element u in $\{\mathtt{00}, \mathtt{01}, \mathtt{12}, \mathtt{20}, \mathtt{21}\}$ the three factors of length 2 of $\phi(u)$ are in $\{\mathtt{00}, \mathtt{01}, \mathtt{12}, \mathtt{20}, \mathtt{21}\}$. $\square$

Lemma A.3.

If w is a factor of $\mathbf{x}$ then $\big||w|_{\mathtt{1}} - |w|_{\mathtt{2}}\big| \leq 1$. In particular, the letters $\mathtt{1}$ and $\mathtt{2}$ alternate in $\mathbf{x}$.

Proof. Let w be a factor of $\mathbf{x}$. There are two cases to consider.

If w can be de-substituted (that is, $w = \phi(v)$ for some v), then $|w|_{\mathtt{1}} = |w|_{\mathtt{2}}$ since $|\phi(i)|_{\mathtt{1}} = |\phi(i)|_{\mathtt{2}}$ for all $i \in \{\mathtt{0}, \mathtt{1}, \mathtt{2}\}$.

If w cannot be de-substituted, then either w has even length and occurs at an odd index in $\mathbf{x}$, or w has odd length. If w has odd length, then deleting either the first or last letter results in a word that can be de-substituted, so $\big||w|_{\mathtt{1}} - |w|_{\mathtt{2}}\big| \leq 1$. If w has even length and occurs at an odd index, then its first letter is $\mathtt{0}$ or $\mathtt{2}$ and its last letter is $\mathtt{0}$ or $\mathtt{1}$; deleting the first and last letters results in a word that can be de-substituted, so $\big||w|_{\mathtt{1}} - |w|_{\mathtt{2}}\big| \leq 1$.

Finally, observe that if for all factors of a word u, the numbers of two letters x and y differ by at most 1, then x and y alternate in u. $\square$

Lemma A.4.

Let τ be the morphism defined by $\tau : \mathtt{0} \mapsto \mathtt{0}, \mathtt{1} \mapsto \mathtt{2}, \mathtt{2} \mapsto \mathtt{1}$. If w is a factor of $\mathbf{x}$, then $\tau(w)^{\mathrm{R}}$ is also a factor of $\mathbf{x}$.

Proof. We first prove by induction that

$$\tau(\phi(\mathtt{2}u\mathtt{1}))^{\mathrm{R}} = \phi(\tau(\mathtt{12}u)^{\mathrm{R}})$$

for every factor of the form $2u1$ of $\mathbf{x}$.

One checks that this is true for 21 and 2001. If $2u1$ is a factor not equal to 21 nor 2001, then u must contain a 2 and we can write $2u1 = 2u'12u''1$ where $2u'1$ and $2u''1$ are factors of $\mathbf{x}$. By the inductive hypothesis we have

$$\begin{aligned}
\tau(\phi(2u1))^{\mathrm{R}} &= \tau(\phi(2u'12u''1))^{\mathrm{R}} \\
&= \tau(\phi(2u''1))^{\mathrm{R}}\tau(\phi(2u'1))^{\mathrm{R}} \\
&= \phi(\tau(12u'')^{\mathrm{R}})\phi(\tau(12u')^{\mathrm{R}}) \\
&= \phi(\tau(12u'12u'')^{\mathrm{R}}) \\
&= \phi(\tau(12u)^{\mathrm{R}}).
\end{aligned}$$

We now prove the lemma by induction on the length of w. One can check by hand that the lemma is true for w of length at most 15. Assume the lemma is true for every factor of length at most $n \geq 15$, and let w be a factor of length $n+1$. Then w is a factor of $\phi(v)$ for some factor v of $\mathbf{x}$ with $\frac{n+1}{2} \leq |v| \leq \frac{n+3}{2}$.

Since all factors of length 4 contain a 1 and a 2, there exists a factor u such that v is a factor of $2u1$ and $|2u1| \leq \frac{n+3}{2} + 6$. In particular, w is a factor of $\phi(2u1)$ and $\tau(w)^{\mathrm{R}}$ is a factor of $\tau(\phi(2u1))^{\mathrm{R}}$. To obtain the conclusion, we just need to show that $\tau(\phi(2u1))^{\mathrm{R}}$ is a factor of $\mathbf{x}$.

As by Lemma A.2, a 2 is always preceded by a 1 in $\mathbf{x}$, the word $12u$ is a factor of $\mathbf{x}$ and it has length $|12u| \leq \frac{n+3}{2} + 6 \leq n$. By inductive hypothesis, $\tau(12u)^{\mathrm{R}}$ is a factor of $\mathbf{x}$. Hence $\phi(\tau(12u)^{\mathrm{R}})$ is also a factor. Finally, using the previous result, $\tau(\phi(2u1))^{\mathrm{R}} = \phi(\tau(12u)^{\mathrm{R}})$ is a factor of $\mathbf{x}$. □

We can now express $\mathcal{P}_{\mathbf{x}}^{(1)}$ in terms of Δ_0.

Proof of Proposition 2.39. Let w be a factor of $\mathbf{x}$ of length $|w| = n$.

If $|w| - |w|_0 = |w|_1 + |w|_2$ is even, it follows from Lemma A.3 that $|w|_1 = |w|_2$. Therefore every factor of length n containing exactly $|w|_0$ zeros is abelian-equivalent to w, so the pair $(n, |w|_0)$ determines a unique abelian equivalence class of factors.

If $|w| - |w|_0$ is odd, then by Lemma A.3 either $|w|_1 = |w|_2 + 1$ or $|w|_2 = |w|_1 + 1$. By Lemma A.4, there exists another factor, $v = \tau(w)^{\mathrm{R}}$, of length n with $|v|_0 = |w|_0$ and $|v|_1 - |v|_2 = |w|_2 - |w|_1$. Therefore both possibilities occur. So the number of abelian equivalence classes corresponding to a pair $(n, |w|_0)$ is 2.

There are $\Delta_0(n) + 1$ possible values for the number of 0's in a factor of length n. Since each value occurs for some factor, we have

$$\begin{aligned}
\mathcal{P}_{\mathbf{x}}^{(1)}(n) &= \sum_{i=\min_0(n)}^{\max_0(n)} \begin{cases} 1 & \text{if } n-i \text{ is even} \\ 2 & \text{if } n-i \text{ is odd} \end{cases} \\
&= \sum_{j=n-\max_0(n)}^{n-\min_0(n)} \begin{cases} 1 & \text{if } j \text{ is even} \\ 2 & \text{if } j \text{ is odd.} \end{cases}
\end{aligned}$$

Therefore $\mathcal{P}_{\mathbf{x}}^{(1)}(n) = \frac{3}{2}\Delta_0(n) + c(n)$, where $c(n)$ depends only on the parities of $\Delta_0(n)$ and $n - \min_0(n)$; computing four explicit values allows one to determine the values of $c(n)$ and obtain the equation claimed for $\mathcal{P}_{\mathbf{x}}^{(1)}(n)$. □

A.3.2 $\Delta_0(n)_{n\geq 0}$ and $(\min_0(n) \bmod 2)_{n\geq 0}$ satisfy recurrence relations

We break the proof of Proposition 2.40 into three parts, covered by Lemmas A.5, A.7 and A.9. We first deal with powers of 2.

Lemma A.5.

Let $\ell \in \mathbb{N}$, $\ell \geq 1$. We have $\mathcal{P}_{\mathbf{x}}^{(1)}(2^\ell) = 4$, $\Delta_0(2^\ell) = 2$,

$$\max_0(2^{\ell+1}) = 2^\ell - \min_0(2^\ell) \text{ and } \min_0(2^{\ell+1}) = 2^\ell - \max_0(2^\ell).$$

Proof. Recall from Chapter 1 that $\Psi(w) = (|w|_{\mathtt{0}}, |w|_{\mathtt{1}}, |w|_{\mathtt{2}})$ is the Parikh vector of w. We show by induction that

$$\begin{aligned}&\{\Psi(w) : w \text{ factor of } \mathbf{x} \text{ with } |w| = 2^\ell\}\\ &\qquad = \{P_\ell + (0,0,0), P_\ell + (-2,1,1), P_\ell + (-1,1,0), P_\ell + (-1,0,1)\}\end{aligned}$$

and that

$$\Psi(\phi^\ell(\mathtt{0})) = \begin{cases} P_\ell & \text{if } \ell \text{ is even} \\ P_\ell + (-2,1,1) & \text{if } \ell \text{ is odd} \end{cases}$$

$$\Psi(\phi^\ell(\mathtt{2})) = \begin{cases} P_\ell + (-2,1,1) & \text{if } \ell \text{ is even} \\ P_\ell & \text{if } \ell \text{ is odd,} \end{cases}$$

where $P_\ell = (\frac{2^\ell+4}{3}, \frac{2^\ell-2}{3}, \frac{2^\ell-2}{3})$ if ℓ is odd and $P_\ell = (\frac{2^\ell+2}{3}, \frac{2^\ell-1}{3}, \frac{2^\ell-1}{3})$ if ℓ is even. Since Parikh vectors of factors of length 2^ℓ can take exactly four values, the conclusion is immediate.

The result is true for $\ell \in \{1, 2\}$. Let $\ell > 2$ and assume the result holds for $\ell - 1$. Let w be a factor of length 2^ℓ.

If w can be de-substituted, then we have $w = \phi(v)$ for some factor v of length $2^{\ell-1}$, and $\Psi(w) = (2|v|_{\mathtt{2}}, |v|_{\mathtt{0}} + |v|_{\mathtt{1}}, |v|_{\mathtt{0}} + |v|_{\mathtt{1}})$. Using the inductive hypothesis, it is easy to check that $\Psi(w) = P_\ell$ or $\Psi(w) = P_\ell + (-2,1,1)$ and that the equalities for $\Psi(\phi^\ell(\mathtt{0})), \Psi(\phi^\ell(\mathtt{2}))$ are satisfied.

If w cannot be de-substituted, then w occurs at an odd index in $\mathbf{x}$ and w is of the form

$$\mathtt{0}^{-1}\phi(v)\mathtt{0}, \quad \mathtt{1}^{-1}\phi(v)\mathtt{1}, \quad \mathtt{0}^{-1}\phi(v)\mathtt{1} \quad \text{or} \quad \mathtt{1}^{-1}\phi(v)\mathtt{0}$$

for some factor v of length $2^{\ell-1}$. If w is of one of the first two forms, then $\Psi(w) = \Psi(\phi(v))$ and $\Psi(w) = P_\ell$ or $\Psi(w) = P_\ell + (-2,1,1)$ (as in the previous case).

If $w = \mathtt{0}^{-1}\phi(v)\mathtt{1}$, then w can also be written as $w = \mathtt{0}\phi(u)\mathtt{2}^{-1}$ for some factor u of length $2^{\ell-1}$. So both Parikh vectors $\Psi(\phi(v))$ and $\Psi(\phi(u))$ belong to $\{P_\ell, P_\ell + (-2,1,1)\}$. Since by construction $\phi(v)$ has two more zeros than $\phi(u)$, we obtain $\Psi(\phi(v)) = P_\ell$ and $\Psi(\phi(u)) = P_\ell + (-2,1,1)$. Thus $\Psi(w) = \Psi(\phi(v)) + (-1,1,0) = P_\ell + (-1,1,0)$.

Similarly, if $w = \mathtt{1}^{-1}\phi(v)\mathtt{0}$, then $\Psi(w) = P_\ell + (-1,0,1)$.

To conclude the proof, we just need to show that these four cases actually occur for all ℓ. Since $\{\Psi(\phi^\ell(\mathtt{0})), \Psi(\phi^\ell(\mathtt{2}))\} = \{P_\ell, P_\ell + (-2,1,1)\}$, consider all factors of length 2^ℓ occurring between two consecutive occurrences of $\Psi(\phi^\ell(\mathtt{0}))$ and $\Psi(\phi^\ell(\mathtt{2}))$. By continuity of the number of $\mathtt{0}$'s, one of these factors must have a Parikh vector equal to $P_\ell + (-1,1,0)$ or $P_\ell + (-1,0,1)$.

Using Lemma A.4, we obtain that w is a factor of length 2^ℓ with $\Psi(w) = P_\ell + (-1, 1, 0)$ if and only if $\tau(w)^{\mathrm{R}}$ is a factor of length 2^ℓ with $\Psi(w) = P_\ell + (-1, 0, 1)$. So all four values actually occur. □

To show Lemmas A.7 and A.9, we first prove the following technical result.

Lemma A.6.

Let u be a factor of $\mathbf{x}$ of length $n \geq 1$. Let $\max_2(n)$ (resp. $\min_2(n)$) denote the maximum (resp. minimum) of $\{|w|_2 : w \text{ factor of } \mathbf{x} \text{ of length } n\}$. We have

- $|u|_2 = \max_2(n)$ if and only if $|\phi(u)|_0 = \max_0(2n)$,
- $|u|_2 = \min_2(n)$ if and only if $|\phi(u)|_0 = \min_0(2n)$.

Proof. For the first assertion, assume that $|u|_2 = \max_2(n)$ and suppose $|\phi(u)|_0 < \max_0(2n)$. Note that $|\phi(u)|_0 = 2|u|_2$ by definition of ϕ. Let v be a factor of length $2n$ such that $|v|_0 = \max_0(2n)$, which is even by Lemma 2.38. In addition, we can assume that v starts with $\mathtt{00}$. Indeed, if it is not the case, then either v starts with $\mathtt{01}$ and ends with $\mathtt{0}$, or v is of the form $t\mathtt{00}s$ where t does not contain any zero. In the first case, we can consider the word $\mathtt{0}v\mathtt{0}^{-1}$ that starts with $\mathtt{00}$ and has $\max_0(2n)$ zeros. In the second case, we can consider the word $\mathtt{00}sw$ for some w with $|w| = |t|$. This factor has also $\max_0(2n)$ zeros. Therefore v can be de-substituted. So $v = \phi(z)$ and $|z|_2 = \frac{1}{2}|v|_0 > |u|_2$, which is a contradiction.

For the other direction, assume $|\phi(u)|_0 = \max_0(2n)$ and suppose $|u|_2$ does not maximize the number of $\mathtt{2}$'s. Then there exists a factor v of length n such that $|v|_2 = \max_2(n)$. Hence,

$$|\phi(v)|_0 = 2|v|_2 > 2|u|_2 = |\phi(u)|_0 = \max_0(2n),$$

which is a contradiction. Similar arguments hold for the second assertion. □

Lemma A.7.

If $\ell \geq 2$ and $0 \leq r \leq 2^{\ell-1}$, then

$$\max_0(2^\ell + r) = \max_0(2^\ell) + \max_0(r),$$
$$\min_0(2^\ell + r) = \min_0(2^\ell) + \min_0(r).$$

Proof. We work by induction on ℓ. One checks the case $\ell = 2$. Let $\ell > 2$ and assume the statements are true for $\ell - 1$. Let $0 \leq r \leq 2^{\ell-1}$.

Assume first that r is even. We exhibit a factor of length $2^\ell{+}r$ that has $\max_0(2^\ell) + \max_0(r)$ zeros and maximizes the number of $\mathtt{0}$'s. By the inductive hypothesis, the result is true for $2^{\ell-1} + r/2$. So there exists a factor u of length $2^{\ell-1} + r/2$ with a number of zeros equal to $\min_0(2^{\ell-1} + r/2) = \min_0(2^{\ell-1}) + \min_0(r/2)$. In addition, we can assume that u maximizes the number of $\mathtt{2}$'s. Indeed, since $|u|_0 = \min_0(2^{\ell-1} + r/2)$, $|u|_1 + |u|_2$ is maximal among all factors of length $2^{\ell-1} + r/2$. If the number of $\mathtt{1}$ and $\mathtt{2}$ in u is even, then $|u|_2 = |u|_1$ is maximal. Otherwise, either $|u|_2 = |u|_1 + 1$ and $|u|_2$ is maximal, or $|u|_2 = |u|_1 - 1$ and u

does not maximize the number of $\mathtt{2}$'s. In the last case, by Lemma A.4, we can consider the factor $\tau(u)^{\mathrm{R}}$ which satisfies $|\tau(u)^{\mathrm{R}}|_{\mathtt{0}} = |u|_{\mathtt{0}}$ and $|\tau(u)^{\mathrm{R}}|_{\mathtt{2}} = |u|_{\mathtt{1}}$. Hence, $\tau(u)^{\mathrm{R}}$ minimizes the number of $\mathtt{0}$'s and maximizes the number of $\mathtt{2}$'s.

Let us write $u = vw$ with $|v| = 2^{\ell-1}$ and $|w| = r/2$. Then, as $|v|_{\mathtt{0}} + |w|_{\mathtt{0}} = |u|_{\mathtt{0}}$ is equal to $\min_0(2^{\ell-1}) + \min_0(r/2)$, the words v and w minimize the number of $\mathtt{0}$'s for words of their respective lengths. The word v maximizes also the number of $\mathtt{2}$'s for factors of length $2^{\ell-1}$ because $|v|$ and $|v|_{\mathtt{0}} = \min_0(2^{\ell-1})$ are even by Lemma A.5 and so is $|v|_{\mathtt{1}} + |v|_{\mathtt{2}}$. Since u maximizes the number of $\mathtt{2}$'s and $|v|_{\mathtt{2}} = |v|_{\mathtt{1}}$, the word w also maximizes the number of $\mathtt{2}$'s. Hence, by Lemma A.6, $\phi(u)$, $\phi(v)$ and $\phi(w)$ maximize the number of $\mathtt{0}$'s for words of their respective lengths. Thus,

$$\max_0(2^\ell + r) = |\phi(u)|_{\mathtt{0}} = |\phi(v)|_{\mathtt{0}} + |\phi(w)|_{\mathtt{0}} = \max_0(2^\ell) + \max_0(r).$$

If r is odd, we still have $0 \le r - 1 \le r + 1 \le 2^{\ell-1}$ and we can use the previous results:

$$\begin{aligned} \max_0(2^\ell + r - 1) &= \max_0(2^\ell) + \max_0(r - 1), \\ \max_0(2^\ell + r + 1) &= \max_0(2^\ell) + \max_0(r + 1). \end{aligned}$$

Note that $\max_0$ is even for even values and can only grow by 0 or 1. So there are two cases to consider: either $\max_0(2^\ell+r+1) = \max_0(2^\ell+r-1)$ or $\max_0(2^\ell+r+1) = \max_0(2^\ell+r-1)+2$.

If the two maxima are equal, then we have $\max_0(r + 1) = \max_0(r - 1)$, $\max_0(2^\ell + r) = \max_0(2^\ell + r - 1)$ and $\max_0(r) = \max_0(r - 1)$, and we are done. Otherwise, the two maxima differ by 2, and then $\max_0(r + 1) = \max_0(r - 1) + 2$, $\max_0(2^\ell + r) = \max_0(2^\ell + r - 1) + 1$ and $\max_0(r) = \max_0(r - 1) + 1$, and we are done.

A similar proof shows that $\min_0(2^\ell + r) = \min_0(2^\ell) + \min_0(r)$. □

Lemma A.9 will follow directly from the following lemma.

Lemma A.8.

If $\ell \ge 2$ and $2^{\ell-1} \le r \le 2^\ell$, then

$$\begin{aligned} \max_0(2^{\ell+1}) &= \max_0(2^\ell + r) + \min_0(2^\ell - r), \\ \min_0(2^{\ell+1}) &= \min_0(2^\ell + r) + \max_0(2^\ell - r). \end{aligned}$$

Moreover, there is a factor of length $2^{\ell+1}$ maximizing (resp. minimizing) the number of $\mathtt{0}$'s such that the prefix of length $2^\ell + r$ also maximizes (resp. minimizes) the number of $\mathtt{0}$'s. In addition, the first equality $\max_0(2^{\ell+1}) = \max_0(2^\ell+r)+\min_0(2^\ell-r)$ holds even if $\ell = 1$.

Proof. We proceed by induction on ℓ. One checks that the results are true for $\ell = 2$ and, for the first equality, for $\ell = 1$. Let $\ell > 2$ and assume both equalities hold for $\ell - 1$. Let $2^{\ell-1} \le r \le 2^\ell$.

Assume first that r is even. By the inductive hypothesis, there exists a factor $u = vw$ of length 2^ℓ such that

$$|u|_{\mathtt{0}} = \min_0(2^\ell) = \min_0(2^{\ell-1} + r/2) + \max_0(2^{\ell-1} - r/2),$$

$|v| = 2^{\ell-1} + r/2$ and v minimizes the number of 0's. Hence, $|v|_0 = \min_0(2^{\ell-1} + r/2)$ and $|w|_0 = \max_0(2^{\ell-1} - r/2)$.

Observe that u maximizes the number of 2's as $|u|$ and $|u|_0 = \min_0(2^\ell)$ are even. In addition, we can assume that v also maximizes the number of 2's. Indeed, if v is of even length, $|v|_0 = \min_0(2^{\ell-1} + r/2)$ implies $|v|_2$ is maximal. If v is of odd length and v does not maximize the number of 2's, then it ends with 1. Thus, v is followed by a 2. In particular, v occurs at an even index in $\mathbf{x}$. So is u and $u12$ or $u00$ is a factor of $\mathbf{x}$. If $u12$ is a factor, then consider, instead of u, $u' = z^{-1}u1$ where z denotes the first letter of u. In that case, the prefix of length $2^{\ell-1} + r/2$ of u' is $z^{-1}v2$. It still minimizes the number of 0's and now maximizes the number of 2's. Assume now that $u00$ is a factor. Observe that $\mathbf{x}$ is the fixed point of ϕ. So it is also the fixed point of ϕ^2. Therefore, $\mathbf{x}$ is a concatenation of blocks of length 4 of the form $\phi^2(0) = \phi^2(1) = 1200$ and $\phi^2(2) = 1212$. Since $u00$ is a factor of $\mathbf{x}$, the only extension of this factor is $12u00$ as $|u| = 2^\ell \equiv 0 \pmod 4$. Consider then $u' = 2u2^{-1}$.

Since $|u|_1 = |u|_2$ and $|v|_2 \geq |v|_1$, $|w|_1 \geq |w|_2$. Thus, as $|w|_0 = \max_0(2^{\ell-1} - r/2)$, w minimizes the number of 2's. By Lemma A.6, we obtain $|\phi(u)|_0 = \max_0(2^{\ell+1})$, $|\phi(v)|_0 = \max_0(2^\ell + r)$, $|\phi(w)|_0 = \min_0(2^\ell - r)$. So

$$\begin{aligned}\max_0(2^{\ell+1}) &= |\phi(u)|_0 = |\phi(v)|_0 + |\phi(w)|_0 \\ &= \max_0(2^\ell + r) + \min_0(2^\ell - r).\end{aligned}$$

We can show similarly that $\min_0(2^{\ell+1}) = \min_0(2^\ell + r) + \max_0(2^\ell - r)$. Note that in this case, we can assume that the factor u with $|u|_0 = \max_0(2^\ell)$, given by the inductive hypothesis, starts with 00 as in the proof of Lemma A.6.

Assume now that r is odd. Then $2^{\ell-1} \leq r-1 < r+1 \leq 2^\ell$ and we can apply the previous result:

$$\begin{aligned}\max_0(2^{\ell+1}) &= \max_0(2^\ell + r - 1) + \min_0(2^\ell - r + 1) \\ &= \max_0(2^\ell + r + 1) + \min_0(2^\ell - r - 1).\end{aligned}$$

Since $\max_0$ is even for even values and can only grow by 0 or 1, there are two cases to consider: either $\max_0(2^\ell + r - 1) = \max_0(2^\ell + r + 1)$ or $\max_0(2^\ell + r - 1) + 2 = \max_0(2^\ell + r + 1)$.

If the two maxima are equal, then $\min_0(2^\ell - r + 1) = \min_0(2^\ell - r - 1) = \min_0(2^\ell - r)$ and $\max_0(2^\ell + r) = \max_0(2^\ell + r - 1)$, and we are done. Otherwise, the two maxima differ by 2, and then $\min_0(2^\ell - r + 1) - 2 = \min_0(2^\ell - r - 1)$. So $\max_0(2^\ell + r) = \max_0(2^\ell + r - 1) + 1$ and $\min_0(2^\ell - r) = \min_0(2^\ell - r + 1) - 1$, and we are done. Using similar argument, we can conclude that $\min_0(2^{\ell+1}) = \min_0(2^\ell + r) + \max_0(2^\ell - r)$.

For the construction of the factors, one can construct them using the factors $\phi(u)$ and $\phi(u')$ given for $r-1$ and $r+1$ in the previous construction. We consider the same two cases as before.

If the maxima are equal, then $\max_0(2^\ell + r) = \max_0(2^\ell + r - 1)$. By construction, $\phi(u)$ has a prefix $\phi(v)$ of length $2^\ell + r - 1$, maximizing the number of 0's. The letter z following the prefix $\phi(v)$ in $\phi(u)$ is not a 0. Otherwise, $\phi(v)0$ would be a factor of length $2^\ell + r$ with $\max_0(2^\ell + r) + 1$ zeros, which is a contradiction. Hence, $\phi(v)z$ is a prefix of length $2^\ell + r$ of $\phi(u)$ that maximizes the number of 0's.

If $\max_0(2^\ell + r - 1) + 2 = \max_0(2^\ell + r + 1)$, then $\max_0(2^\ell + r) = \max_0(2^\ell + r + 1) - 1$. By construction, $\phi(u')$ has a prefix $\phi(v')$ of length $2^\ell + r + 1$, maximizing the number of 0's. This prefix must end with 0. Otherwise, deleting the last letter of $\phi(v')$ would give a factor of length $2^\ell + r$ with $\max_0(2^\ell + r + 1) = \max_0(2^\ell + r) + 1$ zeros, which is a contradiction. Hence, $\phi(v')0^{-1}$ is a prefix of length $2^\ell + r$ of $\phi(u')$ that maximizes the number of 0's.

A similar construction yields a factor of length $2^{\ell+1}$ minimizing the number of `0`'s such that the prefix of length $2^\ell + r$ also minimizes the number of `0`'s. ☐

The previous lemma permits us to reformulate some relations between the two sequences $\max_0(n)_{n\geq 0}$ and $\min_0(n)_{n\geq 0}$.

Lemma A.9.

If $\ell \geq 2$ and $2^{\ell-1} \leq r \leq 2^\ell$, then

$$\begin{aligned}\max\nolimits_0(2^\ell + r) &= 2^\ell - \min\nolimits_0(2^{\ell+1} - r),\\ \min\nolimits_0(2^\ell + r) &= 2^\ell - \max\nolimits_0(2^{\ell+1} - r).\end{aligned}$$

The first equality holds even if $\ell = 1$.

Proof. One can check the first equality for $\ell = 1$. Let $\ell \geq 2$ and $2^{\ell-1} \leq r \leq 2^\ell$. From the previous lemma, we have

$$\max\nolimits_0(2^\ell + r) = \max\nolimits_0(2^{\ell+1}) - \min\nolimits_0(2^\ell - r).$$

Note that, by Lemma A.5, we have $\max_0(2^{\ell+1}) = 2^\ell - \min_0(2^\ell)$. Moreover, by Lemma A.7, since $0 \leq 2^\ell - r \leq 2^\ell$, we get

$$\min\nolimits_0(2^\ell) + \min\nolimits_0(2^\ell - r) = \min\nolimits_0(2^\ell + 2^\ell - r).$$

Since similar relations hold when exchanging $\min_0$ and $\max_0$, the conclusion follows. ☐

The proof of Proposition 2.40 about the reflection relation satisfied by $\Delta_0(n)$ and the recurrence relation of $\min_0(n)$ is now immediate.

Proof of Proposition 2.40. Let $\ell \geq 2$. For $0 \leq r \leq 2^{\ell-1}$, subtracting the two relations provided by Lemma A.7 gives $\Delta_0(2^\ell + r) = \Delta_0(2^\ell) + \Delta_0(r)$. Using the first relation given in Lemma A.5, $\Delta_0(2^\ell) = 2$, it follows that

$$\Delta_0(2^\ell + r) = \Delta_0(r) + 2.$$

Furthermore, $\min_0(2^\ell + r) \equiv \min_0(2^\ell) + \min_0(r) \pmod 2$ by Lemma A.7. Since we have $\min_0(2^\ell) \equiv 0 \pmod 2$ by Lemma A.5, we obtain

$$\min\nolimits_0(2^\ell + r) \equiv \min\nolimits_0(r) \pmod 2.$$

For $2^{\ell-1} < r < 2^\ell$, subtracting the two relations provided by Lemma A.9 permits us to conclude that

$$\Delta_0(2^\ell + r) = \Delta_0(2^{\ell+1} - r).$$

Moreover, using Lemma A.9, we get

$$\begin{aligned}\min\nolimits_0(2^\ell + r) &\equiv \max\nolimits_0(2^{\ell+1} - r) \pmod 2\\ &\equiv \min\nolimits_0(2^{\ell+1} - r) + \Delta_0(2^{\ell+1} - r) \pmod 2.\end{aligned}$$

☐

A.3.3 The 2-abelian complexity of p is piecewise-defined

In this section, we want to compute the 2-abelian complexity of the period-doubling word in terms of the 1-abelian complexity of its 2-block coding, $\mathbf{x} = \text{block}(\mathbf{p}, 2)$. We require several preliminary results.

Proposition A.10.

Let u and v be factors of $\mathbf{p}$ of length n. Let u' and v' be the 2-block codings of u and v. The factors u and v are 2-abelian equivalent if and only if u' and v' are abelian equivalent and either u' and v' both start with 2 or none of them start with 2.

Proof. By Lemma 1.48, u and v are 2-abelian equivalent if and only if they start with the same letter and have the same number of factors 00, 01 and 10. The number of 00 (respectively 01 and 10) in u is exactly the number of 0 (resp. 1 and 2) in u'. Moreover, u starts with 0 (resp. by 1) if and only if u' starts with 0 or 1 (resp. by 2). Therefore, u and v are 2-abelian equivalent if and only if u' and v' are abelian equivalent and both start with 2 or none of them start with 2. □

If the class does not split, we say that it leads to only one class.

Lemma A.11.

Let $\mathcal{X}$ be an abelian equivalence class of factors of length n of $\mathbf{x}$. If the number of 1's in an element of $\mathcal{X}$ differs from the number of 2's, then $\mathcal{X}$ leads to only one 2-abelian equivalence class of $\mathbf{p}$.

Proof. It is enough to prove that if an element of $\mathcal{X}$ starts with 2, all the other elements of $\mathcal{X}$ start with 2. If u starts with 2, then all the elements of $\mathcal{X}$ have more 2's than 1's. But any factor with more 2's than 1's starts with a 2. □

Corollary A.12.

If n is odd, $\mathcal{P}^{(2)}_{\mathbf{p}}(n+1) = \mathcal{P}^{(1)}_{\mathbf{x}}(n)$.

Proof. Let $\mathcal{X}$ be an abelian equivalence class of factors of odd length n. If no element of $\mathcal{X}$ starts with a 2, $\mathcal{X}$ leads to only one 2-abelian equivalence class of factors of $\mathbf{p}$. So assume that there is a factor u in $\mathcal{X}$ starting with 2. Since n is odd, we can write $u = 2\phi(u')$. Then the number of 0's in u is even and there is a different number of 2's than 1's. By Lemma A.11, $\mathcal{X}$ again leads to a unique 2-abelian equivalence class of $\mathbf{p}$. □

Corollary A.13.

Let $\mathcal{X}$ be an abelian equivalence class of factors of $\mathbf{x}$ of even length n with an odd number of zeros. Then $\mathcal{X}$ leads to only one 2-abelian equivalence class of $\mathbf{p}$.

Proof. Factors in $\mathcal{X}$ have an odd number of 1's and 2's counted together, so the number of 1's and the number of 2's are different and we can apply Lemma A.11. $\square$

Thus, an abelian equivalence class $\mathcal{X}$ of factors of length n of $\mathbf{x}$ can possibly lead to two 2-abelian equivalence classes of factors of length $n+1$ of $\mathbf{p}$ only if n is even and if there are an even number of zeros in $\mathcal{X}$. In most cases $\mathcal{X}$ will indeed lead to two different equivalence classes. The exceptions are identified by the following lemma.

Lemma A.14.

Let n be a positive even integer and n_0 such that $\min_0(n) \leq n_0 \leq \max_0(n)$. Let $\mathcal{X}$ be an abelian equivalence class of factors of $\mathbf{x}$ of length n with exactly n_0 zeros.

- We have $n_0 = \max_0(n)$ and $\mathrm{MJ}_0(n) = 1$ if and only if every factor u in $\mathcal{X}$ can be written as $u = 00u'00$.
- We have $n_0 = \min_0(n)$ and $\mathrm{mj}_0(n) = 1$ if and only if every factor u in $\mathcal{X}$ is preceded and followed only by 00.

Proof. We start by proving the first part of the lemma. Assume that all the elements of $\mathcal{X}$ have the form $00u'00$. In particular, n_0 is even. If $n_0 \neq \max_0(n)$, it means that there is a factor v of length n with n_0+1 zeros. Indeed, sliding a window of length n from a word of $\mathcal{X}$ to a factor with $\max_0(n)$ zeros gives factors with all possibilities between n_0 and $\max_0(n)$ for the number of zeros. Since $|v|_0$ is odd and n is even, we must have $v = 0\phi(v')1$ or $v = 2\phi(v')0$. But then $0^{-1}v2$ or $1v0^{-1}$ is an element of $\mathcal{X}$ not of the form $00u'00$, a contradiction. Hence $n_0 = \max_0(n)$. If $\mathrm{MJ}_0(n) = 0$, then $\max_0(n-1) = n_0$ and there is a factor v of odd length $n-1$ with even number n_0 of 0's. We must have $v = 2\phi(v')$ or $v = \phi(v')1$ but then $1v$ or $v2$ is an element of $\mathcal{X}$ not of the form $00u'00$, a contradiction and $\mathrm{MJ}_0(n) = 1$.

For the other direction, assume that $n_0 = \max_0(n)$ and $\mathrm{MJ}_0(n) = 1$. In particular, $\max_0(n-1) = n_0 - 1$. Assume there exists a factor u of $\mathcal{X}$ not of the form $u = 00u'00$. Since u has even length and even number of 0's, we must have $u = 01u'20$ or u has its first or last letter y not equal to 0. In the first case, $v = 001u'$ has length $n-1$ and n_0 zeros, a contradiction. In the second case, removing the letter y leads also to a factor of length $n-1$ with n_0 zeros.

The second part of the lemma is similar. Assume first that all the elements of $\mathcal{X}$ are preceded and followed by 00. In particular, n_0 is even. If $n_0 \neq \min_0(n)$, there is a factor v of length n with $n_0 - 1$ zeros. Since $|v|_0$ is odd but n is even, we must have $v = 0\phi(v')1$ or $v = 2\phi(v')0$ but then $0v1^{-1}$ or $2^{-1}v0$ is an element of $\mathcal{X}$ that starts or ends with 00 and so is preceded or followed by 12, a contradiction. Hence we have $n_0 = \min_0(n)$. If $\mathrm{mj}_0(n) = 0$, then $\min_0(n+1) = n_0$ and there is a factor v of odd length $n+1$ with even number n_0 of 0's. We must have $v = 2\phi(v')$ or $v = \phi(v')1$ but then $\phi(v')$ is an element of $\mathcal{X}$ without a 00 preceding or following it.

For the other direction, assume that $n_0 = \min_0(n)$ and $\mathrm{mj}_0(n) = 1$. In particular, it means that $\min_0(n+1) = n_0 + 1$. If there exists a factor u of $\mathcal{X}$ such that $1u$, $2u$, $u1$ or $u2$ is a factor, then $\min_0(n+1) \leq n_0$, a contradiction. Hence all the factors u of $\mathcal{X}$ can only be extended by $0u0$. Finally, note that $u \in \mathcal{X}$ cannot occur in $\mathbf{x}$ at odd index. In other words, any $u \in \mathcal{X}$ can be de-substituted. Indeed, if it is not the case, then u is of the form $0\phi(u')0$, $0\phi(u')1$, $2\phi(u')0$ or $2\phi(u')1$. If u is of the first form, then $\phi(u')001$ is a factor of length $n+1$ with only n_0 zeros, which is a contradiction. Otherwise, u is of one of the last three forms. Then either $u2$ or $1u$ is a factor of $\mathbf{x}$, which is not possible. So the only extension of u as a factor of $\mathbf{x}$ is $00u00$. □

Lemma A.15.

Let n be a positive even integer and n_0 even such that $\min_0(n) \leq n_0 \leq \max_0(n)$. Let $\mathcal{X}$ be an abelian equivalence class of factors of $\mathbf{x}$ of length n with n_0 zeros. The class $\mathcal{X}$ leads to only one 2-abelian equivalence class of $\mathbf{p}$ if and only if $n_0 = \min_0(n)$ and $\mathrm{mj}_0(n) = 1$ or $n_0 = \max_0(n)$ and $\mathrm{MJ}_0(n) = 1$. Otherwise, $\mathcal{X}$ splits into two classes.

Proof. The factors in $\mathbf{x}$ of length $n = 2$ are $00, 01, 12, 21, 20$. The two classes to consider are $\mathcal{X}_1 = \{00\}$, which leads to one class, and $\mathcal{X}_2 = \{12, 21\}$, which splits into two classes. Since $\mathrm{MJ}_0(2) = 1$ and $\mathrm{mj}_0(2) = 0$, the proposition is true.

Hence let $n \geq 4$ even. If $n_0 = \min_0(n)$ and $\mathrm{mj}_0(n) = 1$, then by Lemma A.14, all the elements of $\mathcal{X}$ are preceded by 00. In particular, they all start with 1 and $\mathcal{X}$ leads to only one 2-abelian equivalence class. Similarly, if $n_0 = \max_0(n)$ and $\mathrm{MJ}_0(n) = 1$, then by Lemma A.14, all the elements of $\mathcal{X}$ start with 0 and we have only one class.

Assume now that $\mathcal{X}$ leads to only one class. If an element u of $\mathcal{X}$ starts with 2, we have $u = 2\phi(u')1$ since n and n_0 are even. Then $1u1^{-1}$ is an element of $\mathcal{X}$ starting with 1 and $\mathcal{X}$ splits into two classes. Hence every element u of $\mathcal{X}$ starts with 0 or 1. Assume there exists a factor u in $\mathcal{X}$ that starts with a 1. Then $u = 12\phi(u')$ and u cannot be followed by a 1 since otherwise $1^{-1}u1$ would be an element of $\mathcal{X}$ starting with 2. Hence u is always followed by 00 and so ends with 12. Similarly, it can only be preceded by 00. Hence all the factors in $\mathcal{X}$ starting with a 1 are preceded and followed by 00. In particular, if a factor in $\mathcal{X}$ starts with 1 and occurs in $\mathbf{x}$ at index i, then the two factors starting at indices $i-1$ and $i+1$ in $\mathbf{x}$ have $n_0 + 1$ zeros. Assume now there exists a factor u in $\mathcal{X}$ starting with a 0. Then, u can be de-substituted. Otherwise, as n and n_0 are even, u is of the form $0\phi(u')0$ where $\phi(u')$ ends with 12. Thus $2\phi(u')2^{-1}$ is an element of $\mathcal{X}$ starting with 2, which is a contradiction. Hence u starts with 00. If u ends with 12, then again, $2u2^{-1}$ is an element of $\mathcal{X}$ starting with 2. Hence $u = 00\phi(u')00$ and all elements of $\mathcal{X}$ starting with 0 start and end with 00. In particular, if a factor in $\mathcal{X}$ starts with 0 and occurs in $\mathbf{x}$ at index i, then the two factors starting at indices $i-1$ and $i+1$ in $\mathbf{x}$ have $n_0 - 1$ zeros.

If no elements of $\mathcal{X}$ start with 1 or no elements start with 0, we are done by Lemma A.14. Otherwise, since one can show that $\mathbf{x}$ is uniformly recurrent, we can assume that there exist a factor $u \in \mathcal{X}$ that starts with 0 and occurs at index i in $\mathbf{x}$, and a factor $v \in \mathcal{X}$ that starts with 1 and occurs at index $i + \ell$ in $\mathbf{x}$, such that any factor w_s of length n occurring at index $i + s$ in $\mathbf{x}$ does not belong to $\mathcal{X}$ for $0 < s < \ell$. Then w_1 has $n_0 - 1$ zeros whereas $w_{\ell-1}$ has $n_0 + 1$ zeros. But there is no factor w_s with n_0 zeros. This is a contradiction since the number of 0's changes by at most one between two factors of the same length starting at consecutive indexes. □

Proof of Proposition 2.44. The case n odd is given by Corollary A.12. Assume now that n is even. Then by Lemma 2.38, $\min_0(n)$ and $\max_0(n)$ are even, and therefore $\Delta_0(n)$ is even as well. Let $\mathcal{X}$ be an abelian equivalence class of factors of $\mathbf{x}$ of length n. Let n_0 be the number of 0's in the elements of $\mathcal{X}$. There are exactly $\frac{\Delta_0(n)}{2}$ odd values of n_0 and $\frac{\Delta_0(n)}{2}+1$ even values. By Corollary A.13, if n_0 is odd, $\mathcal{X}$ leads to one 2-abelian equivalence class of $\mathbf{p}$. By Lemma A.15, $\mathcal{X}$ splits into two classes except for $n_0 = \min_0(n)$ if $\mathrm{mj}_0(n) = 1$ and for $n_0 = \max_0(n)$ if $\mathrm{MJ}_0(n) = 1$. Hence there are in total $\frac{\Delta_0(n)}{2} + 1 - \mathrm{MJ}_0(n) - \mathrm{mj}_0(n)$ cases where $\mathcal{X}$ leads to two 2-abelian equivalence classes of $\mathbf{p}$ instead of one and this is exactly the difference between $\mathcal{P}^{(2)}_{\mathbf{p}}(n+1)$ and $\mathcal{P}^{(1)}_{\mathbf{x}}(n)$. $\square$

Appendix B

Constant 2-labellings of weighted cycles

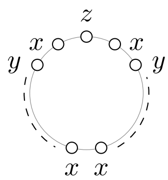
We give in this appendix the proofs omitted in Chapter 6 about constant 2-labellings of weighted cycles.

Contents

B.1 Weighted cycles of Type3mod

Constant 2-labellings of cycles of Type3mod are similar to the ones of cycles of Type1mod.

Lemma B.1.



Let $p > 3$ be an integer such that $p \equiv 3 \pmod 4$. Let $\mathcal{C}_p$ be a cycle of Type3mod, i.e., $z(xy)^{\frac{p-3}{4}}xx(yx)^{\frac{p-3}{4}}$, with $x \neq y$. If c is a non-trivial constant 2-labelling, then $p \equiv 0 \pmod 3$ and c is 3-periodic of pattern period `110`.

Proof. Let $p > 2$ be an integer such that $p \equiv 1 \pmod 4$ and let $\mathcal{C}_p$ be a cycle of Type3mod with $x \neq y$. Assume that c is a non-trivial constant 2-labelling of $\mathcal{C}_p$. The colouring c is not alternate since p is odd. Hence, without loss of generality, we can assume that there exist two consecutive black vertices. Moreover, we can suppose that these vertices are the vertices 0 and 1 of $\mathcal{C}_p$.

For the colouring c, we let α_x, α_y denote respectively the number of black vertices with weight x and y. We have $a = \alpha_x x + \alpha_y y + z$. We consider the colour of the vertex $\frac{p+1}{2}$.

Assume first that $c(\frac{p+1}{2}) = \mathtt{1}$. Then, for the colouring $c \circ \mathcal{R}_1$, the sum of the weights of black vertices is

$$a = (\alpha_x - 2)y + z + x + \alpha_y x + y = (\alpha_y + 2)x + (\alpha_x - 2)y + z$$

since under a 1-rotation, any black vertex with weight y becomes a black vertex of weight x. Similarly, under a 1-rotation, any black vertex with weight x becomes a black vertex of weight y, except for two vertices: the vertex 1 which becomes the vertex with weight z and the vertex $\frac{p+1}{2}$ which becomes the vertex $\frac{p-1}{2}$ with weight x. As the weights x and y are distinct, it implies that $\alpha_x = \alpha_y + 2$. We set $\alpha := \alpha_y$ for a shorter notation.

Let i be the smallest integer in $\{0, \ldots, \frac{p-1}{2} - 1\}$ such that $c(i+1) = \mathtt{0}$ and assume $c(\frac{p+1}{2} + \ell) = \mathtt{1}$ for any $\ell \in \{0, \ldots, i\}$ (otherwise, consider the colouring $c \circ \mathcal{R}_{\frac{p+1}{2}}$ instead of c). Then $c(\frac{p+1}{2} + i + 1) = \mathtt{0}$ as depicted in Figure B.1. With the colouring $c \circ \mathcal{R}_{i+1}$, we obtain a sum of the weights of black vertices equal to $b = (\alpha + 2)x + (\alpha + 1)y$ (Figure B.2). To conclude this case, consider the vertex $i + 2$ and observe that whatever value is assigned to $c(i + 2)$, we obtain a contradiction (Figure B.3).

Therefore, we have $c(\frac{p+1}{2}) = \mathtt{0}$ and $a = \alpha_x x + \alpha_y y + z$ as in the beginning. Observe that the previous reasoning means that for any integer j, we have[1]

$$c \circ \mathcal{R}_j(0) = \mathtt{1} = c \circ \mathcal{R}_j(1) \Rightarrow c \circ \mathcal{R}_j\left(\frac{p+1}{2}\right) = \mathtt{0}. \tag{B.1}$$

With the colouring $c \circ \mathcal{R}_1$, the sum of the weights of black vertices is

$$a = (\alpha_x - 1)y + z + \alpha_y x + x = (\alpha_y + 1)x + (\alpha_x - 1)y + z.$$

Since $x \neq y$, we get $\alpha_x = \alpha_y + 1$. We set $\alpha := \alpha_y$ for a shorter notation.

[1] Equation (B.1) is exactly the same as Equation (6.1) obtained for cycles of Type1mod.

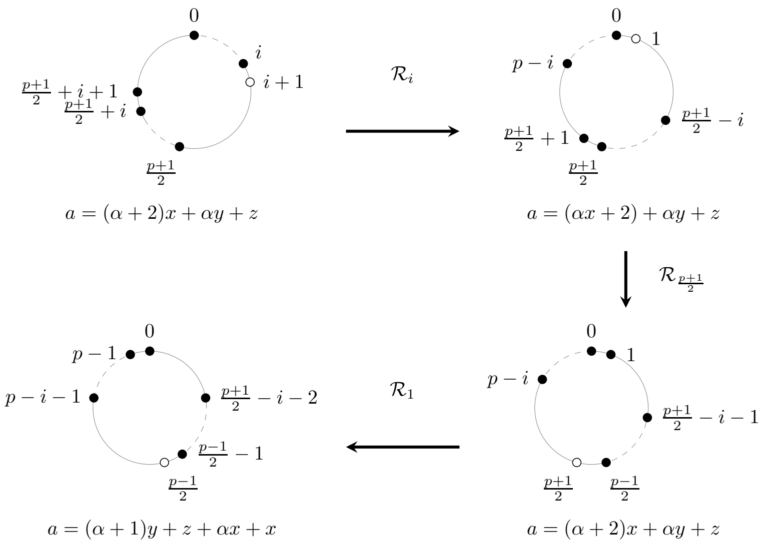


Figure B.1: Rotations of the colouring c of a Type3mod cycle with $c(\frac{p+1}{2}+i+1)=\mathtt{1}$, and their corresponding weighted sums of black vertices which are not all equal.

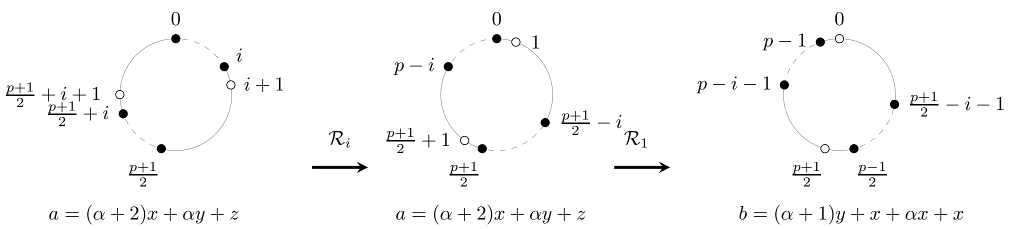


Figure B.2: Rotations of the colouring c of a Type3mod cycle with $c(\frac{p+1}{2}+i+1)=\mathtt{0}$, and their corresponding weighted sums of black vertices.

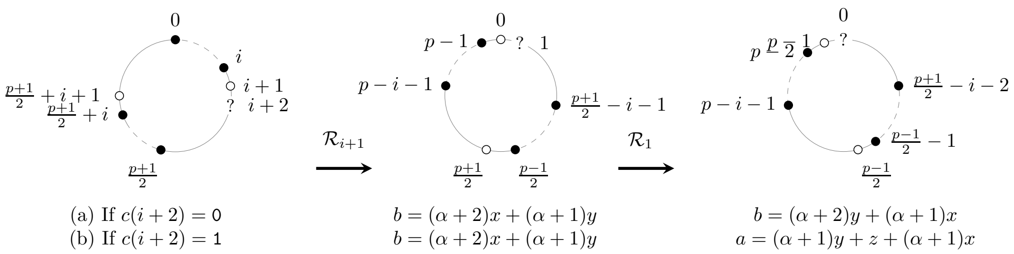


Figure B.3: Rotations of the colouring c of a Type3mod cycle and their corresponding weighted sums of black vertices depending on the colour $c(i+2)$.

Let i be the smallest integer in $\{0, \ldots, \frac{p-1}{2} - 1\}$ such that $c(i+1) = \mathtt{0}$. From Equation (B.1), we have $c(\frac{p+1}{2}+\ell) = \mathtt{0}$ for any $\ell \in \{0, \ldots, i-1\}$. Moreover, we have $c(\frac{p+1}{2}+i) = \mathtt{1}$. Indeed, assume that $c(\frac{p+1}{2}+i) = \mathtt{0}$ (Figure B.4), then with the colouring $c \circ \mathcal{R}_{i+1}$ we obtain a sum of the weights of black vertices equal to $b = (\alpha+1)x + (\alpha+1)y$. As c is a constant 2-labelling, with the colouring $c \circ \mathcal{R}_{\frac{p+1}{2}}$, we have the same weighted sum b. Then it implies that the weighted sum b with the colouring $c \circ \mathcal{R}_{\frac{p+1}{2}+1}$ has a different value, which is a contradiction. So $c(\frac{p+1}{2}+i) = \mathtt{1}$ and with the colouring $c \circ \mathcal{R}_{i+1}$, we have a sum of the weights of black vertices equal to $b = (\alpha+2)x + \alpha y$ (Figure B.5).

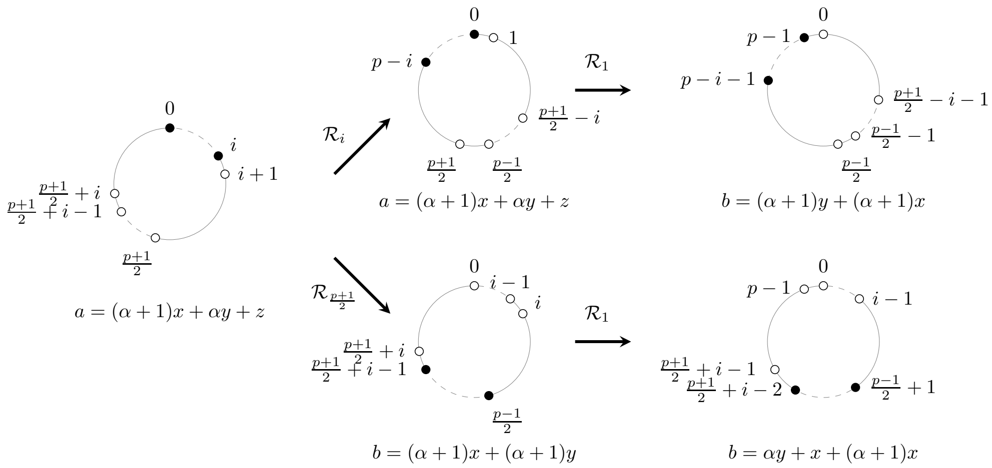


Figure B.4: Rotations of the colouring c of a Type3mod cycle with $c(\frac{p+1}{2}+i) = \mathtt{0}$, and their corresponding weighted sums b of black vertices which are not all equal.

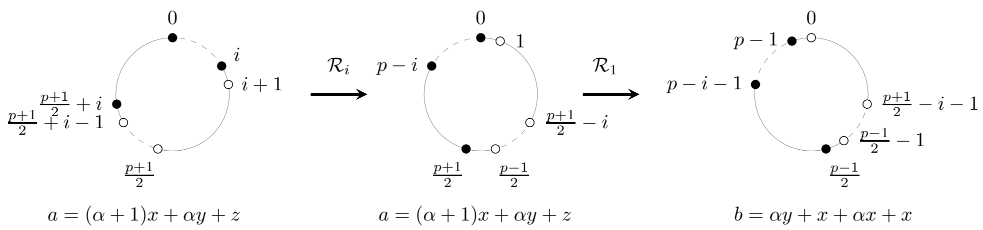


Figure B.5: Rotations of the colouring c of a Type3mod cycle with $c(\frac{p+1}{2}+i) = \mathtt{1}$, and their corresponding weighted sums of black vertices.

From $b = (\alpha+2)x + \alpha y$, it follows that i must be equal to 2, otherwise the colouring $c \circ \mathcal{R}_{\frac{p+1}{2}}$ leads to a different sum of the weights of black vertices (Figure B.6). Then we have $c(3) = \mathtt{1}$ (Figure B.7). Similarly $c(\frac{p+1}{2}+2) = \mathtt{1}$ (Figure B.8).

Therefore, the colouring $c \circ \mathcal{R}_{\frac{p+1}{2}+1}$ has the same configuration as the colouring c, i.e., the vertices $0, 1$ are black and the vertex $\frac{p+1}{2}$ is white. We can apply the same argument as before. Hence, the colouring c must be 3-periodic of pattern period `110` and the number p of vertices is such that $p \equiv 0 \pmod 3$. $\square$

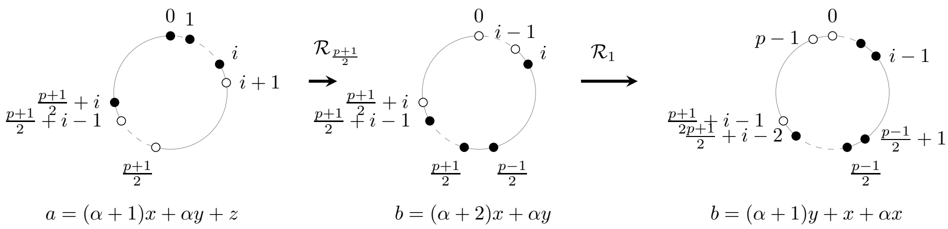


Figure B.6: Rotations of the colouring c of a Type3mod cycle with $c(j) = \mathtt{1}$ for any $0 \leq j \leq i$ with $i > 1$, and their corresponding weighted sums of black vertices distinct which are not all equal.

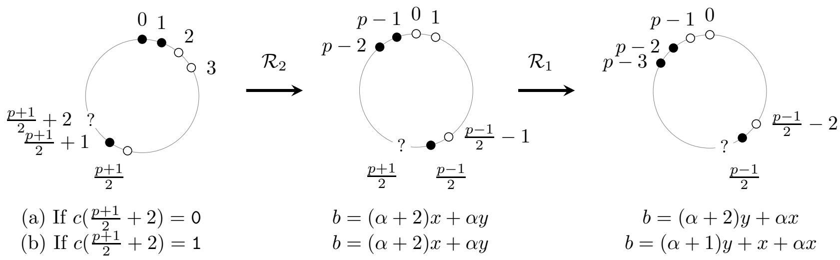


Figure B.7: Rotations of the colouring c of a Type3mod cycle with $c(0) = c(1) = \mathtt{1}$, $c(\frac{p+1}{2}+1) = \mathtt{1}$ and $c(3) = c(\frac{p+1}{2}) = \mathtt{0}$, and their corresponding weighted sums of black vertices depending on the colour $c(\frac{p+1}{2}+2)$.

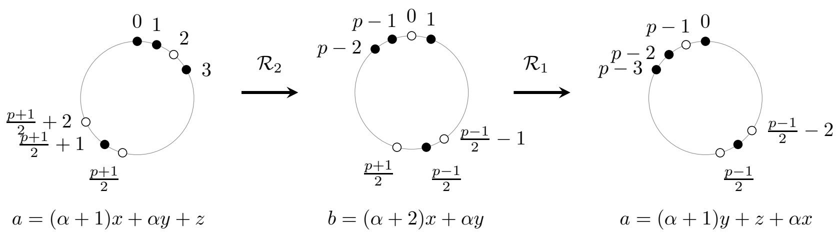
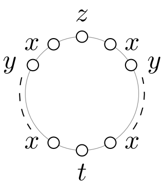


Figure B.8: Rotations of the colouring c of a Type3mod cycle with $c(0) = c(1) = \mathtt{1}$, $c(\frac{p+1}{2}+1) = \mathtt{1}$ and $c(3) = c(\frac{p+1}{2}) = c(\frac{p+1}{2}+2) = \mathtt{0}$, and their corresponding weighted sums of black vertices which are not all equal.

B.2 Weighted cycles of Type4mod

Constant 2-labellings of cycles of Type4mod are similar to the ones of cycles of Type2mod.

Lemma B.2.



Let $p \equiv 4 \pmod 4$ with $p > 4$ and let $\mathcal{C}_p$ be a weighted cycle of Type4mod represented by $z(xy)^{\frac{p-4}{4}}xtx(yx)^{\frac{p-4}{4}}$ where the weights x, y, t are not all equal. If c is a non-trivial constant 2-labelling, then c is one of the following colouring

- alternate,
- $\frac{p}{2}$-anti-periodic,
- $\frac{p}{2}$-periodic if $x = y$; $\frac{p}{2}$-periodic and such that the numbers of black vertices of weight x and y are equal when $c(0) = \mathtt{0}$ if $y \neq x$,
- if $t = \frac{p}{4}x + (1 - \frac{p}{4})y$, c can be moreover such that $c(i) = c(i + \frac{p}{2}) = \mathtt{1}$ for all even $i \in \{0, \dots, \frac{p}{2} - 1\}$ and $c(i) \neq c(i + \frac{p}{2})$ for all odd $i \in \{0, \dots, \frac{p}{2} - 1\}$ (up to a 1-rotation).

Proof. Let $p \equiv 4 \pmod 4$ with $p > 4$ and let $\mathcal{C}_p$ be a weighted cycle of Type4mod represented by $z(xy)^{\frac{p-4}{4}}xtx(yx)^{\frac{p-4}{4}}$ where the weights x, y, t are not all equal. Clearly, the alternate colouring is a constant 2-labelling of $\mathcal{C}_p$ with weighted sums $a = (\frac{p}{2} - 2)y + z + t$ and $b = \frac{p}{2}x$.

The case where the weights t and y are equal follows from Lemma 6.8. Hence, we suppose from now on that $t \neq y$. Consider a non-trivial constant 2-labelling c of $\mathcal{C}_p$ that is not the alternate colouring. Without loss of generality, we may assume that there exist two consecutive vertices black vertices and that they are the vertices 0 and 1. Hence, we assume that $c(0) = c(1) = \mathtt{1}$. We let α_x, α_y and α_t denote respectively the numbers of the vertices with weight x, y and t for the colouring c. We have then $a = \alpha_x x + \alpha_y y + \alpha_t t + z$. We consider four cases depending on the colours of the vertices $\frac{p}{2}$ and $\frac{p}{2} + 1$.

<u>Case 1</u>: Suppose that $c(\frac{p}{2}) = \mathtt{1} = c(\frac{p}{2} + 1)$. It means in particular that $\alpha_t = 1$. Then for the colouring $c \circ \mathcal{R}_1$, the sum of the weights of the black vertices is

$$a = (\alpha_x - 2)y + z + t + \alpha_y x + x + x = (\alpha_y + 2)x + (\alpha_x - 2)y + t + z.$$

If the weights x and y are distinct, then we have $\alpha_x = \alpha_y + 2$ and we set $\alpha := \alpha_y$ for a shorter notation. Otherwise, we denote by β the number $\alpha_x + \alpha_y$ of black vertices of weights $x = y$. We have

$$\begin{cases} a = (\alpha + 2)x + \alpha y + t + z & \text{if } x \neq y \\ a = \beta x + t + z & \text{if } x = y \end{cases}$$

Let i be the smallest integer in $\{0, \dots, \frac{p}{2} - 1\}$ such that $c(i + 1) = \mathtt{0}$ and assume that $c\left(\frac{p}{2} + \ell\right) = \mathtt{1}$ for all $\ell \in \{0, \dots, i\}$ (otherwise, consider the colouring $c \circ \mathcal{R}_{\frac{p}{2}}$ instead of c). Since $t \neq y$, it follows $c(\frac{p}{2} + i + 1) = \mathtt{0}$ as depicted in Figure B.9.

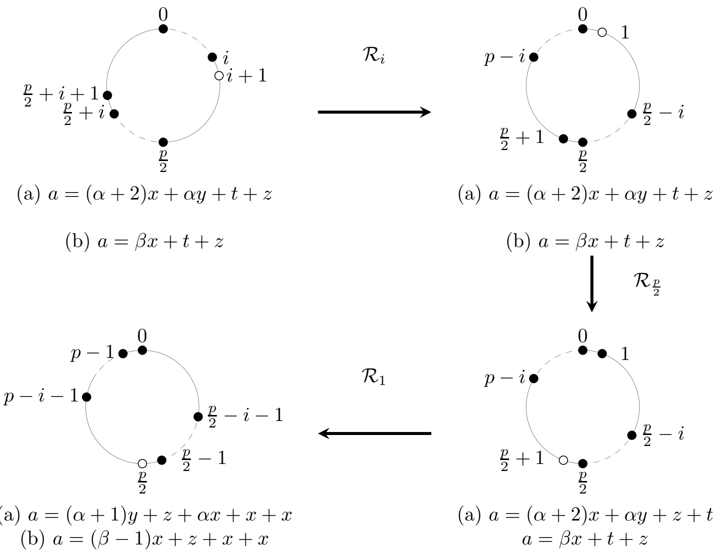


Figure B.9: Rotations of the colouring c of a Type4mod cycle $\mathcal{C}_p$ with $c(\frac{p}{2}+i+1)=\mathtt{1}$, and their corresponding weighted sums of black vertices which are not all equal, where the line (a) corresponds to the case $x \neq y$ and the line (b) to the case $x=y$.

Hence $c(i+1)=c\left(\frac{p}{2}+i+1\right)=\mathtt{0}$. With the colouring $c\circ\mathcal{R}_{i+1}$, we obtain (Figure B.10) a sum of the weights of the black vertices equal to

$$\begin{cases} b=(\alpha+2)(x+y) & \text{if } x\neq y \\ b=(\beta+2)x & \text{if } x=y \end{cases}$$

and the number of black vertices of weight x for the colouring $c\circ\mathcal{R}_{i+1}$ is actually $\alpha+2$ (respectively $\beta+2$) when $x\neq y$ (resp. $x=y$). Observe that if $c(i+2)=\mathtt{0}=c(\frac{p}{2}+i+2)$, then the weighted sum is preserved.

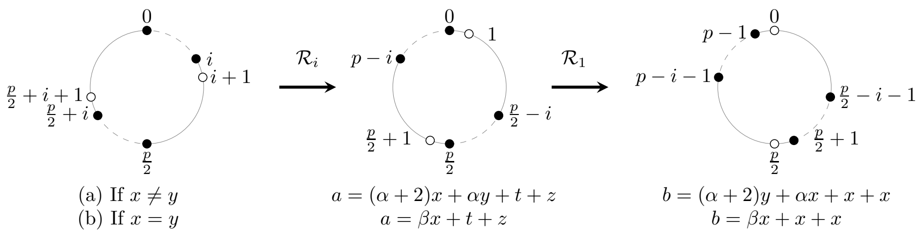


Figure B.10: Rotations of the colouring c of a Type4mod cycle and their corresponding weighted sums of black vertices depending on the equality of the weights x and y.

Let j be the smallest integer in $\{i+1,\dots,\frac{p}{2}-1\}$ such that $c(j+1)=\mathtt{1}$. Without loss of generality, we may assume that $c\left(\frac{p}{2}+\ell\right)=\mathtt{0}$ for all $\ell\in\{i+1,\dots,j\}$. Therefore $c\left(\frac{p}{2}+j+1\right)=\mathtt{1}$, otherwise it implies that $t=y$ which is a contradiction (Figure B.11). Consequently, the sum of the weights of the black vertices with the colouring $c\circ\mathcal{R}_{j+1}$ is $a=(\alpha+2)x+\alpha y+t+z$ (respectively $a=\beta x+t+z$) if the weights x and y are distinct (resp. equal). Moreover, the colourings c and $c \circ \mathcal{R}_{j+1}$ present the same configuration as $c(j+1) = \mathtt{1} = c(\frac{p}{2}+j+1)$ and as the weighted sums are equal. Hence, we can apply the same reasoning given before for c to the colouring $c \circ \mathcal{R}_{j+1}$. It follows that the colouring c is $\frac{p}{2}$-periodic. In particular, we have the following weighted sums

$$\begin{cases} a = (\alpha+2)x + \alpha y + z + t \text{ and } b = (\alpha+1)(x+y) & \text{if } x \neq y \\ a = \beta x + t + z \text{ and } b = (\beta+2)x & \text{if } x = y \end{cases}$$

with β even.

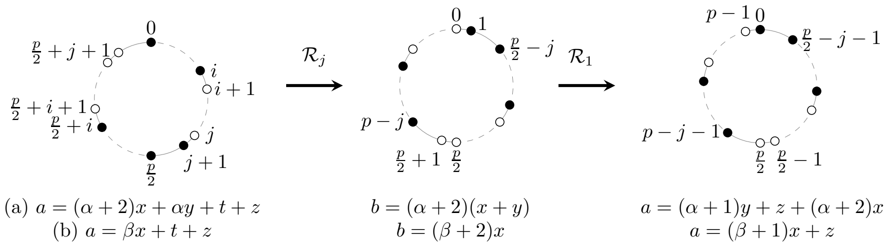


Figure B.11: Rotations of the colouring c of a Type4mod cycle with $c(j+1) \neq c(\frac{p}{2}+j+1) = \mathtt{0}$, and their corresponding weighted sums of black vertices which are not all equal, where the line (a) corresponds to the case $x \neq y$ and the line (b) to the case $x = y$.

<u>Case 2</u>: Suppose that $c(\frac{p}{2}) = \mathtt{0} = c(\frac{p}{2}+1)$. It means that $\alpha_t = 0$ and $a = \alpha_x x + \alpha_y + z$. With the colouring $c \circ \mathcal{R}_1$, the weighted sum of black vertices is equal to

$$a = (\alpha_x - 1)y + z + \alpha_y x + x = (\alpha_y + 1)x + (\alpha_x - 1)y + z$$

as depicted in Figure B.12. Hence, α_x must be equal to $\alpha_y + 1$ if the weights x and y are distinct. In this case, we set $\alpha := \alpha_y$. In the case where $x = y$, we simply set $\beta := \alpha_x + \alpha_y$. We have

$$\begin{cases} a = (\alpha+1)x + \alpha y + z & \text{if } x \neq y \\ a = \beta x + z & \text{if } x = y. \end{cases}$$

We obtain (Figure B.12) the following weighted sum for the colouring $c \circ \mathcal{R}_{\frac{p}{2}}$

$$\begin{cases} b = (\alpha+1)x + \alpha y + t & \text{if } x \neq y \\ b = \beta x + t & \text{if } x = y. \end{cases}$$

Let i be the smallest integer in $\{0, \ldots, \frac{p}{2}-1\}$ such that $c(i+1) = \mathtt{0}$. We may assume that $c\left(\frac{p}{2}+\ell\right) = \mathtt{0}$ for all $\ell \in \{0, \ldots, i\}$. Otherwise, we consider the colouring $c \circ \mathcal{R}_{\frac{p}{2}}$ instead of c and we apply the same reasoning to the complement colouring of c. It follows that $c\left(\frac{p}{2}+i+1\right) = \mathtt{1}$, otherwise we obtain a contradiction as $t \neq y$ (Figure B.13).

Observe that if $c(i+2) = \mathtt{0}$ and $c(\frac{p}{2}+i+2) = \mathtt{1}$, then the weighted sum b for the colouring $c \circ \mathcal{R}_{i+2}$ is preserved. Therefore, let j be the smallest integer in $\{i+1, \ldots, \frac{p}{2}+i\}$ such that $c(j+1) = \mathtt{1}$. Without loss of generality, we suppose that $c\left(\frac{p}{2}+\ell\right) = \mathtt{1}$ for all $\ell \in \{i+1, \ldots, j\}$. It follows that $c\left(\frac{p}{2}+j+1\right) = \mathtt{0}$. Indeed, $c\left(\frac{p}{2}+j+1\right) = \mathtt{1}$ leads to a contradiction as $x \neq t$ (Figure B.14). Therefore, the sum of the weights of the black vertices for the colouring $c \circ \mathcal{R}_{j+1}$ is $a = (\alpha+1)x + \alpha y + z$ (respectively $a = \beta x + z$) if the weights x and y are distinct (resp. equal). Hence, the colourings c and $c \circ \mathcal{R}_{j+1}$ present the same

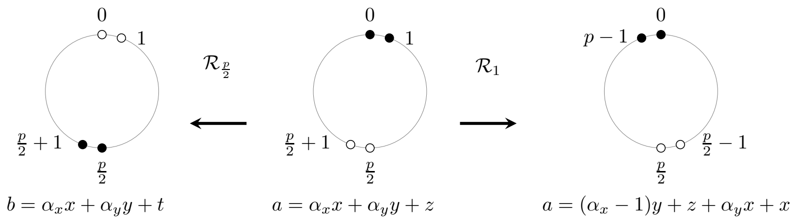


Figure B.12: Rotations of the colouring c of a Type4mod cycle with $c(\frac{p}{2}) = \mathtt{0} = c(\frac{p}{2}+1)$, and their corresponding weighted sums of black vertices.

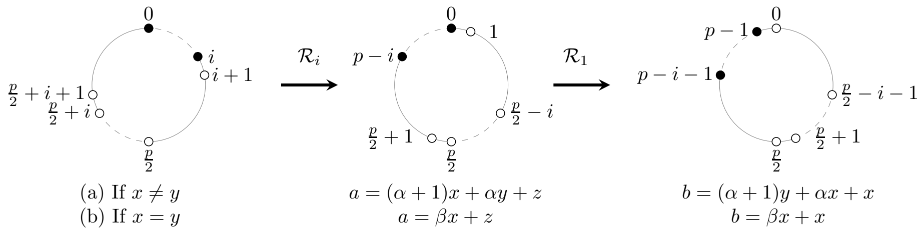


Figure B.13: Rotations of the colouring c of a Type4mod cycle and their corresponding weighted sums of black vertices depending on the equality of the weights x and y.

configuration as $c(j+1) = \mathtt{1}$ and $c(\frac{p}{2}+j+1) = \mathtt{0}$ and as the weighted sums are equal. It follows that c is $\frac{p}{2}$-anti-periodic. In particular, we obtain in this case

$$a = \frac{p}{4}x + \left(\frac{p}{4}-1\right)y + z \text{ and } b = \frac{p}{4}x + \left(\frac{p}{4}-1\right)y + t.$$

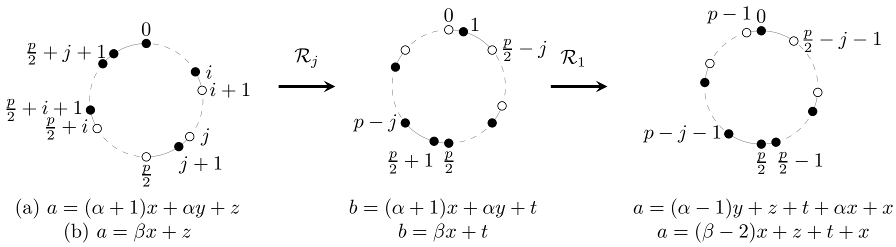


Figure B.14: Rotations of the colouring c of a Type4mod cycle with $c(j+1) \neq c(\frac{p}{2}+j+1) = \mathtt{0}$, and their corresponding weighted sums of black vertices which are not all equal, where the line (a) corresponds to the case $x \neq y$ and the line (b) to the case $x = y$.

<u>Case 3</u>: Suppose that $c(\frac{p}{2}) = \mathtt{0}$ and $c(\frac{p}{2}+1) = \mathtt{1}$. It means that $\alpha_t = 0$ and we have $a = \alpha_x x + \alpha_y y + z$ as the sum of the weights of the black vertices for the colouring c. For the colouring $c \circ \mathcal{R}_1$, we get $a = (\alpha_x - 2)y + z + t + \alpha_y x + x$. Hence,

$$t = (\alpha_x - \alpha_y - 1)x + (\alpha_y - \alpha_x + 2)y. \qquad \text{(B.2)}$$

Moreover, we know the value of the constant b by a $\frac{p}{2}$-rotation. The colouring $c \circ \mathcal{R}_{\frac{p}{2}}$ has weighted sum of black vertices equal to $b = \alpha_x x + \alpha_y y + t$.

Consider the colouring $c \circ \mathcal{R}_1$ which has weighted sum $a = (\alpha_y + 1)x + (\alpha_x - 2)y + z + t$. Then $c(2) \neq c(\frac{p}{2}+2)$. Indeed, the assumption $c(2) = c(\frac{p}{2}+2)$ leads to the following weighted sum

$$\begin{cases} a = (\alpha_y - 1)y + z + t + (\alpha_x - 2)x + 2x & \text{if } c(2) = \mathtt{1} \\ b = (\alpha_y + 1)y + (\alpha_x - 2)x + 2x & \text{if } c(2) = \mathtt{0} \end{cases}$$

for the colouring $c \circ \mathcal{R}_2$. This is a contradiction as $t \neq y$.

If $c(2) = \mathtt{1}$ and $c(\frac{p}{2}+2) = \mathtt{0}$, then the colouring $c \circ \mathcal{R}_2$ has weighted sum $a = \alpha_x x + \alpha_y y + z$ as depicted in Figure B.15. The only possible colours for $c(3)$ and $c(\frac{p}{2}+3)$ are both $\mathtt{1}$. Indeed, if $c(3) = \mathtt{0}$ or $c(\frac{p}{2}+3) = \mathtt{0}$, then we obtain $y = t$ in order to have the weighted sum of the black vertices for the colouring $c \circ \mathcal{R}_3$ equal to a or b (according to the colour $c(3)$). This is a contradiction. Hence, $c(3) = \mathtt{1} = c(\frac{p}{2}+3)$ and we obtain that the weighted sum of the black vertices for the colouring $c \circ \mathcal{R}_3$ is $a = (\alpha_y + 1)x + (\alpha_x - 2)y + t + z$.

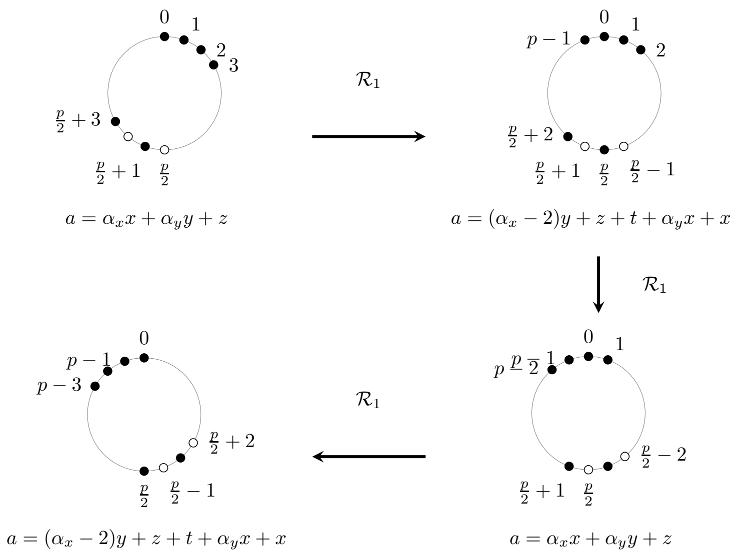


Figure B.15: Rotations of the colouring c of a Type4mod cycle with $c(1) \neq c(\frac{p}{2}+1)$, and their corresponding weighted sums of black vertices.

If $c(2) = \mathtt{0}$ and $c(\frac{p}{2}+2) = \mathtt{1}$, then the colouring $c \circ \mathcal{R}_2$ has weighted sum $b = \alpha_x x + \alpha_y y + t$. Using the same reasoning as above, the only possible colouring of the vertices 3 and $\frac{p}{2}+3$ is $c(3) = \mathtt{1} = c(\frac{p}{2}+3)$. In this case, the weighted sum of the black vertices for the colouring $c \circ \mathcal{R}_3$ is also $a = (\alpha_x - 2)y + z + t + \alpha_y x + x = (\alpha_y + 1)x + (\alpha_x - 2)y + t + z$.

Therefore, for both cases of possible colours $c(2)$ and $c(\frac{p}{2}+2)$, we have

$$c(1) = c(3),\ c\left(\frac{p}{2}+1\right) = c\left(\frac{p}{2}+3\right)$$

and the weighted sums of the black vertices corresponding to the colourings $c \circ \mathcal{R}_1$ and $c \circ \mathcal{R}_3$ are equal to $a = (\alpha_y + 1)x + (\alpha_x - 2)y + z + t$. Hence the colourings $c \circ \mathcal{R}_1$ and $c \circ \mathcal{R}_3$ present the same configuration and we can apply the same reasoning again.

It follows a black pair of diametrically opposed vertices is always followed by a white and black pair and vice versa. That is to say, for any $i \in \{0, \ldots, \frac{p}{2} - 1\}$,

$$\begin{cases} c(i) = c(i + \frac{p}{2}) = \mathtt{1} & \text{if } i \text{ odd} \\ c(i) \neq c(i + \frac{p}{2}) & \text{if } i \text{ even.} \end{cases}$$

Note that this is possible since $p \equiv 0 \pmod 4$. So we have $\alpha_x = \frac{p}{2}$, $\alpha_y = \frac{p}{4} - 1$ and, by Equation (B.2),

$$t = \left(\frac{p}{2} - \frac{p}{4}\right) x + \left(\frac{p}{4} - \frac{p}{2} + 1\right) y = \frac{p}{4}x + \left(1 - \frac{p}{4}\right) y.$$

So we obtain $a = \frac{p}{2}x + (\frac{p}{4} - 1)y + z$ and $b = \frac{3p}{4}x$. By Proposition 6.2, we get $a = (\frac{p}{4} - 1)y + z$ and $b = \frac{p}{4}x$ for the complementary colouring since

$$\sum_{u \in \{0,\dots,p-1\}} w(u) = z + t + \frac{p}{2}x + \left(\frac{p}{2} - 2\right) y.$$

Case 4: Suppose that $c(\frac{p}{2}) = \mathtt{1}$ and $c(\frac{p}{2} + 1) = \mathtt{0}$. This case is similar to Case 3 (Figure B.16) by axial symmetry where the axis is the diameter passing through the vertex 0 and by rotation $\mathcal{R}_{-1}$. Hence we obtain the same conclusion.

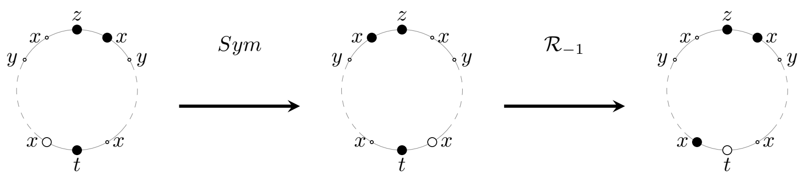
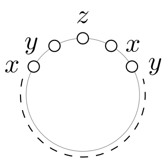


Figure B.16: Axial symmetry *Sym* and (-1)-rotation $\mathcal{R}_{-1}$ of the colouring c of a Type4mod cycle $\mathcal{C}_p$.

□

B.3 Other weighted cycles

Lemma B.3.



Let p be a positive integer. For cycles $\mathcal{C}_p$ represented by $z(xy)^{\frac{p-1}{2}}$ with $x \neq y$, only monochromatic colourings are constant 2-labellings.

Proof. Consider a non-trivial constant 2-labelling c of the cycle represented by $z(xy)^{\frac{p-1}{2}}$ with $x \neq y$ As the number of vertices is odd, c is not the alternate colouring. Without loss of generality, we assume that $c(0) = c(1) = \mathtt{1}$. Let α_x, α_y respectively denote the number of black vertices with weight x and weight y. We have $a = \alpha_x x + \alpha_y y + z$. With the colouring $c \circ \mathcal{R}_1$, we obtain $a = (\alpha_x - 1)y + z + \alpha_y x + y$. So $\alpha_x = \alpha_y$ as $x \neq y$ and we set $\alpha := \alpha_x$ Let i be the smallest integer in $\{0, \dots, p-2\}$ such that $c(i+1) = \mathtt{0}$. With the colouring $c \circ \mathcal{R}_{i+1}$, we obtain a sum of the weights of the black vertices equal to $b = \alpha y + \alpha x + y = \alpha x + (\alpha + 1)y$ (Figure B.17).

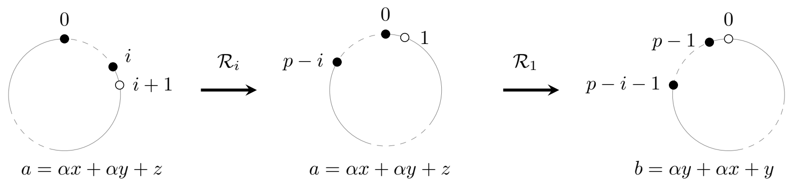


Figure B.17: Rotations of the colouring c and their corresponding weighted sum.

If $c(i+2) = \mathtt{0}$, then we get the weighted sum $b = \alpha_x y + (\alpha_x + 1)x$ with the colouring $c \circ \mathcal{R}_{i+2}$, which is a contradiction. If $c(i+2) = \mathtt{0}$, we obtain the weighted sum $a = (\alpha - 1)y + z + (\alpha + 1)x$, which is a contradiction. Hence, only trivial colourings are constant 2-labellings of this cycle. $\square$

Bibliography

[AB98] Pascal Alessandri and Valérie Berthé. Three distance theorems and combinatorics on words. *Enseign. Math. (2)*, 44(1-2):103–132, 1998.

[Ada03] Boris Adamczewski. Balances for fixed points of primitive substitutions. *Theoret. Comput. Sci.*, 307:47–75, 2003.

[AS92] Jean-Paul Allouche and Jeffrey Shallit. The ring of k-regular sequences. *Theoret. Comput. Sci.*, 98(2):163–197, 1992.

[AS99] Jean-Paul Allouche and Jeffrey Shallit. The ubiquitous Prouhet–Thue–Morse sequence. In *Sequences and their applications (Singapore, 1998)*, Springer Ser. Discrete Math. Theor. Comput. Sci., pages 1–16. Springer, London, 1999.

[AS03a] Jean-Paul Allouche and Jeffrey Shallit. *Automatic sequences*. Cambridge University Press, Cambridge, 2003. Theory, applications, generalizations.

[AS03b] Jean-Paul Allouche and Jeffrey Shallit. The ring of k-regular sequences. II. *Theoret. Comput. Sci.*, 307(1):3–29, 2003. Words.

[Axe03] Maria Axenovich. On multiple coverings of the infinite rectangular grid with balls of constant radius. *Discrete Math.*, 268(1-3):31–48, 2003.

[Bab80] László Babai. On the complexity of canonical labeling of strongly regular graphs. *SIAM J. Comput.*, 9(1):212–216, 1980.

[Bab81] László Babai. On the order of uniprimitive permutation groups. *Ann. of Math. (2)*, 113(3):553–568, 1981.

[Bai13a] Robert Bailey. The metric dimension of small distance-regular and strongly regular graphs, 2013. preprint.

[Bai13b] Robert Bailey. On the metric dimension of imprimitive distance-regular graphs, 2013. preprint.

[BC11] Robert Bailey and Peter Cameron. Base size, metric dimension and other invariants of groups and graphs. *Bull. Lond. Math. Soc.*, 43(2):209–242, 2011.

[BCFR09] Jason Bell, Émilie Charlier, Aviezri Fraenkel, and Michel Rigo. A decision problem for ultimately periodic sets in non-standard numeration systems. *Internat. J. Algebra Comput.*, 19(6):809–839, 2009.

[BCHL04] Nathalie Bertrand, Irène Charon, Olivier Hudry, and Antoine Lobstein. Identifying and locating-dominating codes on chains and cycles. *European J. Combin.*, 25(7):969–987, 2004.

[BD] Valérie Berthé and Vincent Delecroix. s-adic expansions: a combinatorial arithmetic and geometric approach. To appear in RIMS Lecture note 'Kokyuroku Bessatu'.

[BHL00] Uri Blass, Iiro Honkala, and Simon Litsyn. On binary codes for identification. *J. Combin. Des.*, 8(2):151–156, 2000.

[Big73] Norman Biggs. Perfect codes in graphs. *J. Combinatorial Theory Ser. B*, 15:289–296, 1973.

[Bos85] Michael Boshernitzan. A condition for minimal interval exchange maps to be uniquely ergodic. *Duke Math. J.*, 52(3):723–752, 1985.

[BR10] Valérie Berthé and Michel Rigo, editors. *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2010.

[BR11] Jean Berstel and Christophe Reutenauer. *Noncommutative rational series with applications*, volume 137 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2011.

[Brl89] Srečko Brlek. Enumeration of factors in the Thue–Morse word. *Discrete Appl. Math.*, 24(1-3):83–96, 1989. First Montreal Conference on Combinatorics and Computer Science, 1987.

[BSCRF14] Francine Blanchet-Sadri, James Currie, Narad Rampersad, and Nathan Fox. Abelian complexity of fixed point of morphism $0 \mapsto 012, 1 \mapsto 02, 2 \mapsto 1$. *INTEGERS*, 14(#A11), 2014.

[Cam75] Peter Cameron. Partial quadrangles. *Quart. J. Math. Oxford Ser. (2)*, 26:61–73, 1975.

[CD10] Arturo Carpi and Valerio D'Alonzo. On factors of synchronized sequences. *Theoret. Comput. Sci.*, 411(44-46):3932–3937, 2010.

[CH73] Ethan Coven and Gustav Hedlund. Sequences with minimal block growth. *Math. Systems Theory*, 7:138–153, 1973.

[CHL03] Irène Charon, Olivier Hudry, and Antoine Lobstein. Minimizing the size of an identifying or locating-dominating code in a graph is NP-hard. *Theoret. Comput. Sci.*, 290(3):2109–2120, 2003.

[CHLL97] Gérard Cohen, Iiro Honkala, Simon Litsyn, and Antoine Lobstein. *Covering codes*, volume 54 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, 1997.

[CHLM95] Gérard Cohen, Iiro Honkala, Simon Litsyn, and Harold Mattson, Jr. Weighted coverings and packings. *IEEE Trans. Inform. Theory*, 41(6, part 2):1856–1867, 1995.

[CHPZ14] Émilie Charlier, Tero Harju, Svetlana Puzynina, and Luca Zamboni. Abelian bordered factors and periodicity, September 2014. 15e Journées Montoises d'Informatique Théorique.

[CN10] Julien Cassaigne and François Nicolas. Factor complexity. In *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.*, pages 163–247. Cambridge Univ. Press, Cambridge, 2010.

[Cob72] Alan Cobham. Uniform tag sequences. *Math. Systems Theory*, 6:164–192, 1972.

[CRS12] Émilie Charlier, Narad Rampersad, and Jeffrey Shallit. Enumeration and decidable properties of automatic sequences. *Internat. J. Found. Comput. Sci.*, 23(5):1035–1066, 2012.

[Dek79] Frederik Michel Dekking. Strongly non-repetitive sequences and progression-free sets. *J. Combin. Theory Ser. A*, 27(2):181–185, 1979.

[DGHM09] Paul Dorbec, Sylvain Gravier, Iiro Honkala, and Michel Mollard. Weighted codes in Lee metrics. *Des. Codes Cryptogr.*, 52(2):209–218, 2009.

[DGS76] Manfred Denker, Christian Grillenberger, and Karl Sigmund. *Ergodic theory on compact spaces.* Lecture Notes in Mathematics, Vol. 527. Springer-Verlag, Berlin-New York, 1976.

[Die10] Reinhard Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics.* Springer, Heidelberg, fourth edition, 2010.

[dL97] Aldo de Luca. Sturmian words: structure, combinatorics, and their arithmetics. *Theoret. Comput. Sci.*, 183(1):45–82, 1997.

[dLV88] Aldo de Luca and Stefano Varricchio. On the factors of the Thue–Morse word on three symbols. *Inform. Process. Lett.*, 27(6):281–285, 1988.

[Dor07] Paul Dorbec. *Empilements et recouvrements.* PhD thesis, Institut Fourier, 2007.

[Dur98] Fabien Durand. A characterization of substitutive sequences using return words. *Discrete Math.*, 179(1-3):89–101, 1998.

[Eil74] Samuel Eilenberg. *Automata, languages, and machines. Vol. A.* Academic Press, New York, 1974. Pure and Applied Mathematics, Vol. 58.

[EJLR08] Geoffrey Exoo, Ville Junnila, Tero Laihonen, and Sanna Ranto. Locating vertices using codes. In *Proceedings of the Thirty-Ninth Southeastern International Conference on Combinatorics, Graph Theory and Computing*, volume 191, pages 143–159, 2008.

[ELR08] Geoffrey Exoo, Tero Laihonen, and Sanna Ranto. New bounds on binary identifying codes. *Discrete Appl. Math.*, 156(12):2250–2263, 2008.

[Erd61] Paul Erdős. Some unsolved problems. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 6:221–254, 1961.

[FGK$^+$11] Florent Foucaud, Eleonora Guerrini, Matjaž Kovše, Reza Naserasr, Aline Parreau, and Petru Valicov. Extremal graphs for the identifying code problem. *European J. Combin.*, 32(4):628–638, 2011.

[Fis06] Stéphane Fischler. Palindromic prefixes and episturmian words. *J. Combin. Theory Ser. A*, 113(7):1281–1304, 2006.

[FLL+13] Gabriele Fici, Alessio Langiu, Thierry Lecroq, Arnaud Lefebvre, Filippo Mignosi, and Élise Prieur-Gaston. Abelian repetitions in Sturmian words. In *Developments in language theory*, volume 7907 of *Lecture Notes in Comput. Sci.*, pages 227–238. Springer, Heidelberg, 2013.

[FM04] Gašper Fijavž and Bojan Mohar. Rigidity and separation indices of Paley graphs. *Discrete Math.*, 289(1-3):157–161, 2004.

[Fru39] Roberto Frucht. Herstellung von Graphen mit vorgegebener abstrakter Gruppe. *Compositio Math.*, 6:239–250, 1939.

[FXW12] Min Feng, Min Xu, and Kaishun Wang. Identifying codes of lexicographic product of graphs. *Electron. J. Combin.*, 19(4):Paper 56, 2012.

[GJLR14] Sylvain Gravier, Svante Janson, Tero Laihonen, and Sanna Ranto. Graphs where every k-subset of vertices is an identifying set. *Discrete Math. Theor. Comput. Sci.*, 16(1):73–88, 2014.

[GKM08] Sylvain Gravier, Ralf Klasing, and Julien Moncel. Hardness results and approximation algorithms for identifying codes and locating-dominating codes in graphs. *Algorithmic Oper. Res.*, 3(1):43–50, 2008.

[GM07] Sylvain Gravier and Julien Moncel. On graphs having a $V \setminus \{x\}$ set as an identifying code. *Discrete Math.*, 307(3-5):432–434, 2007.

[GMP99] Sylvain Gravier, Michel Mollard, and Charles Payan. Variations on tilings in the Manhattan metric. *Geom. Dedicata*, 76(3):265–273, 1999.

[GMS06] Sylvain Gravier, Julien Moncel, and Ahmed Semri. Identifying codes of cycles. *European J. Combin.*, 27(5):767–776, 2006.

[GMS08] Sylvain Gravier, Julien Moncel, and Ahmed Semri. Identifying codes of Cartesian product of two cliques of the same size. *Electron. J. Combin.*, 15(1):Note 4, 2008.

[GPR+] Sylvain Gravier, Aline Parreau, Sara Rottey, Leo Storme, and Élise Vandomme. Identifying codes in vertex-transitive graphs. preprint.

[GPV13] Sylvain Gravier, Aline Parreau, and Élise Vandomme. Linear formulation of identifying codes in graphs, 2013. Poster presentation at the Fourth PhD-Day of the Belgian Mathematical Society.

[Gre] Florian Greinecker. On the 2-abelian complexity of Thue–Morse subwords.

[GV] Sylvain Gravier and Élise Vandomme. Constant 2-labellings and an application to (r, a, b)-covering codes. preprint.

[GW68] Solomon W. Golomb and Lloyd R. Welch. Algebraic coding and the Lee metric. In *Error Correcting Codes (Proc. Sympos. Math. Res. Center, Madison, Wis., 1968)*, pages 175–194. John Wiley, New York, 1968.

[GW70] Solomon Golomb and Lloyd Welch. Perfect codes in the Lee metric and the packing of polyominoes. *SIAM J. Appl. Math.*, 18:302–317, 1970.

[GW13] Wayne Goddard and Kirsti Wash. ID codes in Cartesian products of cliques. *J. Combin. Math. Combin. Comput.*, 85:97–106, 2013.

[Hig71] Donald Higman. Partial geometries, generalized quadrangles and strongly regular graphs. In *Atti del Convegno di Geometria Combinatoria e sue Applicazioni (Univ. Perugia, Perugia, 1970)*, pages 263–293. Ist. Mat., Univ. Perugia, Perugia, 1971.

[Hig74] Donald Higman. Invariant relations, coherent configurations and generalized polygons. In *Combinatorics (Proc. Advanced Study Inst., Breukelen, 1974), Part 3: Combinatorial group theory*, pages 27–43. Math. Centre Tracts, No. 57. Math. Centrum, Amsterdam, 1974.

[HKS12] Mari Huova, Juhani Karhumäki, and Aleksi Saarela. Problems in between words and abelian words: k-abelian avoidability. *Theoritical Computer Science*, 454:172–177, 2012.

[HL02] Iiro Honkala and Antoine Lobstein. On identifying codes in binary Hamming spaces. *J. Combin. Theory Ser. A*, 99(2):232–243, 2002.

[Hon86] Juha Honkala. A decision method for the recognizability of sets defined by number systems. *RAIRO Inform. Théor. Appl.*, 20(4):395–403, 1986.

[HT91] James Hirschfeld and Joseph Thas. *General Galois geometries*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 1991. Oxford Science Publications.

[HW08] Godfrey Harold Hardy and Edward Maitland Wright. *An introduction to the theory of numbers*. Oxford University Press, Oxford, sixth edition, 2008.

[HZ99] Charles Holton and Luca Zamboni. Descendants of primitive substitutions. *Theory Comput. Syst.*, 32(2):133–157, 1999.

[JL12] Ville Junnila and Tero Laihonen. Optimal identifying codes in cycles and paths. *Graphs Combin.*, 28(4):469–481, 2012.

[JV00] Jacques Justin and Laurent Vuillon. Return words in Sturmian and episturmian words. *Theor. Inform. Appl.*, 34(5):343–356, 2000.

[Kar80] Juhani Karhumäki. Generalized Parikh mappings and homomorphisms. *Inform. and Control*, 47(3):155–165, 1980.

[KCČ+08] Jozef Kratica, Dragoš Cvetković, Mirjana Čangalović, Vera Kovačević-Vujčić, and Jelena Kojić. The metric dimension of strongly regular graphs. In *Proceedings of the XXXV SYMOPIS conference (Belgrade)*, pages 341–344, 2008.

[KCL98] Mark Karpovsky, Krishnendu Chakrabarty, and Lev B. Levitin. On a new class of codes for identifying vertices in graphs. *IEEE Trans. Inform. Theory*, 44(2):599–611, 1998.

[Ker92] Veikko Keränen. Abelian squares are avoidable on 4 letters. In *Automata, languages and programming (Vienna, 1992)*, volume 623 of *Lecture Notes in Comput. Sci.*, pages 41–52. Springer, Berlin, 1992.

[KÖ04] Petteri Kaski and Patric Östergård. The Steiner triple systems of order 19. *Math. Comp.*, 73(248):2075–2092 (electronic), 2004.

[Kra88] J. Kratochvíl. Perfect codes in general graphs. In *Combinatorics (Eger, 1987)*, volume 52 of *Colloq. Math. Soc. János Bolyai*, pages 357–364. North-Holland, Amsterdam, 1988.

[KSZ] Juhani Karhumäki, Aleksi Saarela, and Luca Zamboni. Variations of the Morse–Hedlund theorem for k-abelian equivalence.

[KSZ13] Juhani Karhumäki, Aleksi Saarela, and Luca Zamboni. On a generalization of Abelian equivalence and complexity of infinite words. *J. Combin. Theory Ser. A*, 120(8):2189–2206, 2013.

[Lot97] M. Lothaire. *Combinatorics on words*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1997.

[Lot02] M. Lothaire. *Algebraic combinatorics on words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2002.

[Lov75] László Lovász. On the ratio of optimal integral and fractional covers. *Discrete Math.*, 13(4):383–390, 1975.

[LRRV12] Anne Lacroix, Narad Rampersad, Michel Rigo, and Élise Vandomme. Syntactic complexity of ultimately periodic sets of integers and application to a decision procedure. *Fund. Inform.*, 116(1-4):175–187, 2012.

[Mon06] Julien Moncel. On graphs on n vertices having an identifying code of cardinality $\lceil \log_2(n+1) \rceil$. *Discrete Appl. Math.*, 154(14):2032–2039, 2006.

[Mos96] Brigitte Mossé. Reconnaissabilité des substitutions et complexité des suites automatiques. *Bull. Soc. Math. France*, 124(2):329–346, 1996.

[MP13] Zuzana Masáková and Edita Pelantová. Enumerating abelian returns to prefixes of Sturmian words. In *Combinatorics on words*, volume 8079 of *Lecture Notes in Comput. Sci.*, pages 193–204. Springer, Heidelberg, 2013.

[MR13] Blake Madill and Narad Rampersad. The abelian complexity of the paperfolding word. *Discrete Math.*, 313(7):831–838, 2013.

[MSS11] Johannes Morgenbesser, Jeffrey Shallit, and Thomas Stoll. Thue–Morse at multiples of an integer. *J. Number Theory*, 131(8):1498–1512, 2011.

[OF] The OEIS Foundation. The on-line encyclopedia of integer sequences.

[PF02] N. Pytheas Fogg. *Substitutions in Dynamics, Arithmetics and Combinatorics*, volume 1749 of *Lecture Notes in Mathematics*. Springer-Verlag, 2002.

[PRRV] Aline Parreau, Michel Rigo, Eric Rowland, and Élise Vandomme. A new approach to the 2-regularity of the ℓ-abelian complexity of 2-automatic sequences. preprint.

[PT84] Stanley Payne and Joseph Thas. *Finite generalized quadrangles*, volume 110 of *Research Notes in Mathematics*. Pitman (Advanced Publishing Program), Boston, MA, 1984.

[Puz04] S. A. Puzynina. Periodicity of perfect colorings of an infinite rectangular grid. *Diskretn. Anal. Issled. Oper. Ser. 1*, 11(1):79–92, 2004.

[Puz08] Svetlana Puzynina. Perfect colorings of radius $r > 1$ of the infinite rectangular grid. *Sib. Èlektron. Mat. Izv.*, 5:283–292, 2008.

[PZ13] Svetlana Puzynina and Luca Zamboni. Abelian returns in Sturmian words. *J. Combin. Theory Ser. A*, 120(2):390–408, 2013.

[Rao] Michaël Rao. On some generalizations of abelian power avoidability. Preprint.

[Rig14] Michel Rigo. *Formal Languages, Automata and Numeration Systems 1*, volume 1 of *Network and Telecomunication Series*. Wiley-ISTE, 2014.

[Row10] Eric Rowland. Regularsequences, April 2010.

[RRS13] Narad Rampersad, Michel Rigo, and Pavel Salimov. On the number of abelian bordered words. In *Developments in Language Theory - 17th International Conference, DLT 2013, Marne-la-Vallée, France, June 18-21, 2013. Proceedings*, volume 7907 of *Lecture Notes in Comput. Sci.*, pages 420–432. Springer, Heidelberg, 2013.

[RRS14] Narad Rampersad, Michel Rigo, and Pavel Salimov. A note on abelian returns in rotation words. *Theoret. Comput. Sci.*, 528:101–107, 2014.

[RSTU04] Saikat Ray, David Starobinski, Ari Trachtenberg, and Rachanee Ungrangsi. Robust location detection with sensor networks. *IEEE Journal on Selected Areas in Communications*, 22(6):1016–1025, 2004.

[RSV13] Michel Rigo, Pavel Salimov, and Élise Vandomme. Some properties of abelian return words. *J. Integer Seq.*, 16(2):Article 13.2.5, 21, 2013.

[RSZ11] Gwénaël Richomme, Kalle Saari, and Luca Zamboni. Abelian complexity of minimal subshifts. *J. Lond. Math. Soc. (2)*, 83(1):79–95, 2011.

[RV11] Michel Rigo and Élise Vandomme. Syntactic complexity of ultimately periodic sets of integers. In *Proceedings of the fifth conference LATA (Languages and Automata, Theory and Applications)*, volume 6638 of *Lecture Notes in Comput. Sci.*, pages 477–488. Springer Verlag, 2011.

[RV12] Michel Rigo and Élise Vandomme. 2-abelian complexity of the Thue–Morse sequence, December 2012. Talk presented at Representing Streams I.

[RW14] Douglas Rall and Kirsti Wash. Identifying codes of the direct product of two cliques. *European J. Combin.*, 36:159–171, 2014.

[Sal10] Pavel Salimov. On uniform recurrence of a direct product. *Discrete Math. Theor. Comput. Sci.*, 12(4):1–8, 2010.

[Sla75] Peter Slater. Leaves of trees. In *Proceedings of the Sixth Southeastern Conference on Combinatorics, Graph Theory, and Computing (Florida Atlantic Univ., Boca Raton, Fla., 1975)*, pages 549–559. Congressus Numerantium, No. XIV. Utilitas Math., Winnipeg, Man., 1975.

[Sla87] Peter Slater. Domination and location in acyclic graphs. *Networks*, 17(1):55–64, 1987.

[Sla88] Peter Slater. Dominating and reference sets in a graph. *J. Math. Phys. Sci.*, 22(4):445–455, 1988.

[Tel94] Jan Arne Telle. Complexity of domination-type problems in graphs. *Nordic J. Comput.*, 1(1):157–171, 1994.

[Thu06] Axel Thue. Über unendliche zeichenreihen. *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania*, 7:1–22, 1906.

[Thu12] Axel Thue. Über die gegenseitige Lage gleicher Teile gewisser zeichenreihen. *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania*, 10:1–67, 1912.

[UTS04] Rachanee Ungrangsi, Ari Trachtenberg, and David Starobinski. An implementation of indoor location detection systems based on identifying codes. In *Proceedings of Intelligence in Communication Systems, INTELLCOMM 2004*, volume 3283 of *LNCS*, pages 175–189, 2004.

[Vui01] Laurent Vuillon. A characterization of Sturmian words by return words. *European J. Combin.*, 22(2):263–275, 2001.

[XTH08] Min Xu, Krishnaiyan Thulasiraman, and Xiao-Dong Hu. Identifying codes of cycles with odd orders. *European J. Combin.*, 29(7):1717–1720, 2008.

List of symbols

Arabic letters

$i, j, k, \ell, \ldots$	positive integers
$\mathrm{rep}_k(n)$	base-k representation of n
$\mathrm{srg}(n, k, \lambda, \mu)$	strongly regular graphs on n vertices of degree k with parameters λ, μ
$Aut(G)$	set of all automorphisms of the graph G
$B_r(v)$	ball of radius r with center v
$E(G)$	edge set of the graph G
$\mathrm{Fac}(\mathbf{w})$	set of factors of the infinite word $\mathbf{w}$
$\mathrm{GQ}(s, t)$	generalized quadrangle with parameters s and t
K_n	clique of size n
$N(u)$	open neighbourhood of the vertex u
$N[u]$	closed neighbourhood of the vertex u
$\mathrm{PG}(n, q)$	n-dimensional projective space over $\mathbb{F}_q$
$\mathrm{Pref}(\mathbf{w})$	set of prefixes of the infinite word $\mathbf{w}$
R_α	rotation of angle α
$St(\alpha, \rho)$	Sturmian word with slope α and intercept ρ
$V(G)$	vertex set of the graph G
$\mathcal{APR}_{\mathbf{x},u}$	set of abelian returns to the prefix u of the infinite word $\mathbf{x}$
$\mathcal{APR}_{\mathbf{x}}$	set of all abelian returns to prefixes of $\mathbf{x}$
$\mathcal{AR}_{\mathbf{x},u}$	set of abelian returns to the factor u of $\mathbf{x}$
$\mathcal{AR}_{\mathbf{x}}$	set of all abelian returns to factors of $\mathbf{x}$
$\mathcal{C}$	1-dimensional torus $\mathbb{R}/\mathbb{Z}$ identified with the interval $[0, 1)$
$\mathcal{C}_p$	cycle with p vertices
$\mathcal{D}_u(\mathbf{x})$	derived sequence of $\mathbf{x}$ with respect to u
$\mathcal{E}_u(\mathbf{x})$	abelian derived sequence of $\mathbf{x}$ with respect to u

$\mathcal{E}(G)$	the set of hyperedges of the hypergraph G
$\mathcal{H}_q$	hypercube of dimension q
$\mathcal{K}_k(\mathbf{s})$	k-kernel of the sequence $\mathbf{s} = (s_n)_{n\geq 0}$
$\mathcal{P}_p$	path with p vertices
$\mathcal{P}_{\mathbf{w}}^{\ell}$	ℓ-abelian complexity of the infinite word $\mathbf{w}$
$\mathcal{P}_{\mathbf{w}}^{\infty}$	factor complexity of the infinite word $\mathbf{w}$
$\mathcal{R}_{\mathbf{x},u}$	set of return words to u of $\mathbf{x}$
$\mathbf{f}$	Fibonacci word
$\mathbf{p}$	period-doubling word
$\mathbf{t}$	Thue–Morse word
$\mathbf{x}$	2-block coding of the period-doubling word
$\mathbf{y}$	2-block coding of the Thue–Morse word
$\mathbb{F}_q$	finite field of order q

Greek letters

δ	transition function of an automata
$\beta(G)$	metric dimension of the graph G
$\gamma^{\mathrm{ID}}(G)$	identifying number of the graph G
$\gamma_f^{\mathrm{ID}}(G)$	fractional identifying number of the graph G
$\gamma^{\mathrm{LD}}(G)$	locating-dominating number of the graph G
η	morphism generating the 2-block coding of the period-doubling word
$\mu_{\mathbf{x},u}$	abelian derivation
$\theta_{\mathbf{x},u}$	inverse map of $\mu_{\mathbf{x},u}$
ν	morphism generating the 2-block coding of the Thue–Morse word
σ	morphism generating the Thue–Morse word
τ	a coding
	(In Chapter 2, the coding changing `1` into `2` and vice versa)
τ'	the coding changing `0` into `3` and vice versa
φ	morphism generating the Fibonacci word
ϕ	Golden mean
ψ	morphism generating the period-doubling word
Δ	symmetric difference between sets

Δ_0	difference between the maximal number of $\mathtt{0}$'s and the minimal number of $\mathtt{0}$'s in factors of a given length of $\mathbf{x}$
Δ_{12}	difference between the maximal number of $\mathtt{1}$'s and $\mathtt{2}$'s together and the minimal number of $\mathtt{1}$'s and $\mathtt{2}$'s together in factors of a given length of $\mathbf{y}$
$\Lambda_{\mathbf{x},u}$	derivation map
$\Psi(u)$	Parikh vector of a word u
$\Psi_2(u)$	vector of $\mathbb{N}^{10}$ associated with a word over $\{\mathtt{0},\mathtt{1}\}$

Miscellaneous

$\sim_{ab}$	abelian equivalence between words
$\sim_{ab,\ell}$	ℓ-abelian equivalence between words
$\sim$	equivalence between vectors

Index