UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

EXTENDED SYMMETRIES IN SUBSHIFTS UNDER STRONG RIGIDITY CONDITIONS

TESIS PARA OPTAR AL GRADO DE DOCTOR EN CIENCIAS DE LA INGENIERÍA,
MENCIÓN MODELACIÓN MATEMÁTICA

ALVARO MATÍAS BUSTOS GAJARDO

PROFESOR GUÍA:
MICHAEL HEINRICH SCHRAUDNER

MIEMBROS DE LA COMISIÓN:
MICHAEL BAAKE
SEBASTIÁN DONOSO FUENTES
ALEJANDRO MAASS SEPÚLVEDA
SERVET MARTÍNEZ AGUILERA

Este trabajo ha sido parcialmente financiado por CMM ANID PIA AFB170001 y la beca ANID-PFCHA/Doctorado Nacional/2017-21171061, junto con CRC 1283 (Universität Bielefeld) y UCH-1566.

SANTIAGO DE CHILE
2021

RESUMEN DE LA MEMORIA PARA OPTAR
AL GRADO DE DOCTOR EN CIENCIAS DE LA INGENIERÍA,
MENCIÓN MODELACIÓN MATEMÁTICA
POR: ALVARO MATÍAS BUSTOS GAJARDO
FECHA: 2021
PROF. GUÍA: SR. MICHAEL HEINRICH SCHRAUDNER

EXTENDED SYMMETRIES IN SUBSHIFTS UNDER STRONG RIGIDITY CONDITIONS

Versión en español

En este trabajo de tesis estudiamos los grupos de automorfismos y de simetrías extendidas de sistemas dinámicos simbólicos (en particular, espacios de shift) que presentan una estructura jerárquica de algún tipo. Esto incluye shifts de tipo finito aperiódicos (como el shift de Robinson), shifts sustitutivos multidimensionales y shifts de interés en la teoría de números como el shift libre de cuadrados, que son definidos a partir de una secuencia creciente de reticulados. Estos grupos se estudian desde un punto de vista esencialmente geométrico (en el caso multidimensional), buscando caracterizaciones de éstos en cada contexto mediante herramientas del álgebra, el análisis y la combinatoria.

English version

In this thesis work we study the automorphism and extended symmetry groups of symbolic dynamical systems (in particular, shift spaces), focusing on those that exhibit some kind of hierarchical structure. They include aperiodic shifts of finite type (like the Robinson shift), multidimensional substitutive subshifts and shift spaces of number-theoretical interest, such as the square-free shift, which are defined by an increasing sequence of lattices. These groups are studied from an essentially geometric viewpoint (in the multidimensional case), looking for characterizations in each context via tools from algebra, analysis and combinatorics.

You are filled with
***determination**.*

A Vásquez Castillo

Agradecimientos

Hay demasiada gente a la que quisiera agradecer en estas pocas líneas, y es muy difícil hacerlo en este espacio tan limitado. Estos han sido cuatro años y medio de aprendizaje, una etapa de mi vida que se cierra con la entrega de esta tesis, y en estos años he compartido con mucha gente que me ha ayudado a crecer y formarme, haciéndome alguien muy distinto a quien entró al doctorado el 2017.

Quisiera agradecer en primer lugar a mi familia, que me ha acompañado y tenido infinita paciencia todo este tiempo. A mis padres, Sonia y Daniel, hermanos Pablo y Daniela, sobrinos Ágata y Lautaro y «madrastra» Leonor, que siempre han estado ahí presentes con infinita paciencia, y me han incentivado a cumplir mis metas, y sobre todo a mi hermano Joan, al que extraño cada día. A mis tías, Ester («Chela») y Gladys, que siempre me recibieron con cariño y afecto. También agradezco mucho a mi esposa Javiera, siempre motivándome a dar lo mejor de mí y levantarme cuando mi motivación caía. Gracias por seguir este camino juntos.

Agradezco a mis amigos y amigas que me escucharon y soportaron cuando necesitaba ventilar mis frustraciones porque me faltaba el último paso de la demostración y no me salía, comenzando por mis compañeros de oficina Hugo, Matías y Laura. Echaré mucho de menos los UNOs cuando teníamos una ventana libre, y las juntas en el open space para tratar de entender los tree shifts. También a mis otros (ex-)compañeros del doctorado y pregrado con los que compartíamos tantas risas, penas y frustraciones, como Javiera F., Roberto M., Ariel P., Eduardo S., Javier R., Sebastián T., Felipe M. y tantos más. También, doy gracias a mis alumnos de Cálculo Diferencial, EDO y los otros cursos en los que trabajé en estos cuatro años (y a Francisco, Petra y Nicolás, los auxiliares que tuve en cada ramo). Fue un honor (y muy entretenido) enseñarles lo poco que sabía, y aprendí mucho en el proceso.

Como también hay vida fuera de la universidad, quiero agradecer los momentos de esparcimiento para alejar mi mente de los problemas de la vida (y pucha que ha hecho falta esto en el último año y medio). Gracias a Sebastián I., Dani F. y Emilio por nuestras aventuras juntos. Gracias a Roberto B. («Bob») y Belén por siempre preocuparse de que no estuviera muriendo y de que saliera de mi zona de confort. Agradezco a Justin, Thamara y Mario por esas noches de juegos de mesa y reír juntos, que extraño mucho en estos momentos. Gracias también a Nicolás («Chaparrón»), David («Birlera»), Antonia, Steven («Yasamito»), Paola, Álvaro C., Carlos S., José Luis, Andrés («Ten Rose»), Borja y Camilo, y tantos otros con los que hemos hablado de la vida, del mundo y de nada, y compartido estos cuatro años que han sido de muchos cambios para todos, a escala mundial (por qué será, digo yo. . .).

Encarecidamente le doy las gracias a mi profesor guía Michael Schraudner, quien me guió y acompañó todo este tiempo, se preocupó de enseñarme buenas prácticas y formarme como investigador. También agradezco al profesor Michael Baake, de la Universität Bielefeld, que ha sido un gran apoyo desde mi pasantía en la ciudad de Alemania y desde entonces, siempre dispuesto a ayudarme en mi formación aún después del término de mi estadía. También agradezco a Neil Mañibo y Daniel Luz, a quienes también conocí en Bielefeld. La cuarentena habría sido mucho más pesada sin nuestras reuniones semanales de trabajo y conversación, y las risas ayudaron mucho.

Un abrazo para todos y todas. Los quiero mucho.

Table of Contents

Index of Illustrations

Introduction

Symbolic dynamical systems, which include subshifts, odometers, Bratteli-Vershik systems, and many others, are a subject of great interest in the literature, both of interest by themselves (see, e.g., applications to coding and information theory, Markov chains, etc.), and as representations of other dynamical systems (e.g. the relationship between Sturmian subshifts and circle rotations, Markov partitions for dynamical systems on the interval, etc.). Thus, it is an interesting question to determine whether there exist mathematical objects (*invariants*) that allow us to distinguish between these kinds of subshifts. Another related, interesting question is to see if such an object reflects some aspect of the internal structure of the symbolic space, in an analogous way as how, e.g. homology groups reflect properties about the shape of a surface, such as its Euler characteristic, its "number of holes" and so on.

One of the most studied algebraic invariants for symbolic systems (and topological dynamical systems in general) is the **automorphism group** $\mathrm{Aut}(X, G)$, which consists of the set of all bijections of the phase space to itself which preserve the underlying group action. By this definition, elements from this group must be maps that preserve other dynamical features of the phase space, e.g. automorphisms map asymptotic pairs to asymptotic pairs, periodic or transitive orbits to periodic or transitive orbits, isolated points to isolated points and so on.

In our main case of interest, the dynamical system under scrutiny is a shift space, and the corresponding group action is the action by translations (the **shift action** σ) where every symbol is moved in a fixed direction from its original position. By this nature, automorphism groups are thus described in a combinatorial way, with the well-known **Curtis–Hedlund–Lyndon theorem** being a central part of the theory. Thus, the study of automorphisms is a confluence of dynamics, combinatorics, and algebra, and one may expect that the group under scrutiny reflects some properties of the shift space: "simple" subshifts should have "small" automorphism groups (in a sense given by its algebraic structure), while "complicated" shift spaces should result in more complex groups.

While the automorphism groups of several examples of shift spaces have been thoroughly studied, their analysis is in general complicated, see e.g. the characterization of this group for mixing shifts (and, in particular, the full shift) by Boyle, Lind and Rudolph [19, 62]. Determining whether the automorphism groups of $\{0,1\}^{\mathbb{Z}}$ and $\{0,1,2\}^{\mathbb{Z}}$ are isomorphic or not is still an open problem; this is an example of how this problem is more nuanced than expected. Another example comes from number theory: the two-dimensional **shift of visible points**, which is generated by the indicator function of the set $V \subset \mathbb{Z}^2$ of all points (m, n) with integer coordinates satisfying $\gcd(m, n) = 1$. This shift space, which is discussed in Chapter 6 and has been thoroughly studied before (see e.g. [3] and references therein), has several interesting properties, among them one of the most notorious (for our current purposes) being that it has positive entropy, which intuitively means it is a "complicated" subshift.

However, its automorphism group is actually as close to trivial as possible, being comprised only of the maps induced by the $\mathbb{Z}^2$-group action itself. This shows that "complicatedness" does not always translate to the automorphism group. Furthermore, we may interpret this as a statement about the structure of the points of the shift space itself, and consequently of the set V: there are no locally-defined transformations that preserve the local structure of V beside translations.

However, it is not hard to see that there are many other transformations one may apply to the set V that preserve its structure. For instance, if we think of V as a set of points in the plane, we see that rotations by multiples of $\frac{1}{2}\pi$ and reflections along the coordinate axes leave the set as-is; furthermore, it is not hard to prove that for any invertible integral matrix $A \in \mathrm{GL}_2(\mathbb{Z})$, we have $A \cdot V = V$. One may extend this line of reasoning to the associated shift space X_V itself: if we think of the elements of X_V as indicator functions $\mathbb{1}_W$ of some $W \subset \mathbb{Z}^2$, we have that $\mathbb{1}_W \in \mathsf{X}_V$ if, and only if, $\mathbb{1}_{A\cdot W} \in \mathsf{X}_V$. It is, thus, easy to wonder whether there is some mathematical object which captures this additional structure, for which the automorphism group has proven itself inadequate.

There are several other algebraic invariants one may look upon to, such as the full group $[\![\sigma]\!]$, or the set of all homeomorphisms $\mathsf{X}_V \to \mathsf{X}_V$, which may reflect this structure. However, most of them may be too broad or hard to handle, or lose some other information about the shift space, e.g., by not being as closely related to its dynamical properties such as asymptotic pairs, minimal equicontinuous factor and so on. Thus, it might be a good idea to limit our scope to something that behaves close enough to automorphisms, so as to preserve structures in a similar way. For this, we examine what automorphisms are in an algebraic sense, for a $\mathbb{Z}^d$-shift space (similar considerations apply, of course, for shifts defined over more general group and more diverse dynamical systems), where their structure is given by the equality:

$$f \in \mathrm{Aut}(X, \mathbb{Z}^d) \iff (\forall \boldsymbol{n} \in \mathbb{Z}^d)\colon f \circ \sigma_{\boldsymbol{n}} = \sigma_{\boldsymbol{n}} \circ f, \quad \text{or equivalently} \quad f \circ \sigma_{\boldsymbol{n}} \circ f^{-1} = \sigma_{\boldsymbol{n}},$$

which, in algebraic terms, means that f belongs to the **centralizer** of the subgroup $\langle\sigma\rangle := \{\sigma_{\boldsymbol{n}} : \boldsymbol{n} \in \mathbb{Z}^d\}$ in the set $\mathrm{Homeo}(X)$ of all homeomorphisms of the space X, that is, the set of all homeomorphisms which commute with every shift map $\sigma_{\boldsymbol{n}}$. One may relax this condition so as to apply to the whole set $\langle\sigma\rangle$ instead of to individual elements, that is, $f \circ \langle\sigma\rangle \circ f^{-1} = \langle\sigma\rangle$. That is, we are looking for homeomorphisms $f\colon X \to X$ that satisfy the weaker condition:

$$(\forall \boldsymbol{n} \in \mathbb{Z}^d)(\exists \boldsymbol{m} \in \mathbb{Z}^d)\colon f \circ \sigma_{\boldsymbol{n}} = \sigma_{\boldsymbol{m}} \circ f. \tag{1}$$

While this condition is weaker than the one defining an automorphism, it is still strong enough to preserve characteristic features of a dynamical system, like periodic points and asymptotic pairs. We call the set of all these mappings the **extended symmetry group** $\mathrm{Sym}(X, \mathbb{Z}^d)$ of the space X, and this will be the main subject of study along the publications that compose this work. As we can easily verify, the previous definition is equivalent to saying that $\mathrm{Sym}(X, \mathbb{Z}^d)$ is the **normalizer** of the set $\langle\sigma\rangle$, and is thus the largest subgroup of $\mathrm{Homeo}(X)$ where $\langle\sigma\rangle$ is normal.

It can be proven that $\mathrm{Aut}(X, \mathbb{Z}^d)$ is itself normal in $\mathrm{Sym}(X, \mathbb{Z}^d)$, and that the relation between $\boldsymbol{n}$ and $\boldsymbol{m}$ in (1) is linear, that is, there is some invertible integral matrix $A \in \mathrm{GL}_d(\mathbb{Z})$, depending only on f and not on $\boldsymbol{n}$, such that $\boldsymbol{m} = A\boldsymbol{n}$. This shows that the quotient $\mathrm{Sym}(X, \mathbb{Z}^d)/\operatorname{Aut}(X, \mathbb{Z}^d)$ is some subgroup of $\mathrm{GL}_d(\mathbb{Z})$. The interesting part is that we can give a geometrical interpretation of this quotient, as symmetries *in a geometrical sense* (isometries, affine transformations, etc.) visible in the points of the subshift. For instance, going back to our example of the set of visible points, given some $A \in \mathrm{GL}_d(\mathbb{Z})$, the map $\Phi_A\colon \mathbb{1}_W \mapsto \mathbb{1}_{A\cdot W}$ is always an extended symmetry under this definition, as it can very easily verified to be continuous, with inverse $\Phi_{A^{-1}} = \Phi_A^{-1}$ and satisfying $\Phi_A \circ \sigma_{\boldsymbol{n}} = \sigma_{A\boldsymbol{n}} \circ \Phi_A$. Thus, this new group captures all these geometric symmetries that were “missed” by the usual automorphism group. A similar situation happens with the Robinson tiling, as seen in Chapter 4, where extended symmetries corresponding to rotations by multiples of $\frac{1}{2}\pi$ and reflections along the coordinate axes appear; furthermore, it may be proven that, up to a shift, no other automorphisms or extended symmetries appear, and thus this group is entirely dictated by the geometric structure imposed by the grid of interlocked squares characteristic of this shift space.

This work has as its goal to study extended symmetries from this geometric viewpoint, while also delving in the algebraic and dynamical aspects of the study of this group. It is divided as follows:

- We summarize the prerrequisites on each topic in Chapters 1 (group theory), 2 (ring theory and algebraic number theory) and 3 (symbolic dynamics). This is not intended as a thorough exposition, but as an accessible quick reference or “thesaurus”. While Chapters 1 and 3 are relevant for this whole work, Chapter 1 has been written with the topics dealt with in Chapter 5 in mind first and foremost. Similarly, Chapter 2 is mostly relevant for the topics discussed in Chapter 6, which diverges somewhat from the exposition on the previous two chapters.
- Chapters 4 and 5 deal with automorphisms and extended symmetries in bijective substitutions.
    - Chapter 4 is an adaptation of *Extended symmetry groups for multidimensional subshifts with hierarchical structure* [22], and deals with the multidimensional case from a geometrical and combinatorial viewpoint; it also applies similar techniques to the study of the Robinson tiling.
    - Chapter 5, which is a collaboration with Daniel Luz and Neil Mañibo from Universität Bielefeld, corresponds to a version of *Admissible Reversing and Extended Symmetries for Bijective Substitutions* [23] with some additional commentary, comments both on the one- and multidimensional situations, but pays more attention to the algebraic aspects of these groups, while delving into some algorithms and criteria to determine the group $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)$ from the underlying substitution.
- Chapter 6 is a commented version of *Number-theoretic positive entropy shifts with small centraliser and large normaliser* [4], a collaboration with Michael Baake, Christian Huck, Marius Lemańczyk and Andreas Nickel. This deals with shifts of number-theoretical origin, particularly k-free and $\mathcal{B}$-free shift spaces, and analogous subshifts defined via factorization properties in a ring of algebraic integers, which are studied using the extended symmetry group as a main tool, and connections between this group and the underlying ring’s divisibility properties are studied.

Part I

Preliminaries

Chapter 1

Elementary group theory

This chapter intends to introduce the basics of group theory that will be required as background for this work. As this is intended as a basic reference only, most proofs are omitted or given very brief overviews.

1.1. Basic notions

Although we assume the corresponding definitions to be known by the reader, we shall summarize them below for quick reference. We direct the interested reader to any standard reference book on abstract algebra, such as Lang [68], Hungerford [53] or Grilliet [48], or alternatively specialized books on the subject such as Hall [49].

Definition 1.1 *A* ***group*** *is an ordered pair* $(G, *)$*, where* G *is any set and* $*$*, the* ***group operation*** *is a function*[1] $G \times G \to G$*, satisfying the following properties:*

- ***Associativity:*** $x * (y * z) = (x * y) * z$.
- ***Neutral element:*** *there exists a unique* $1_G \in G$ *such that, for all* $x \in G$*,* $1_G * x = x * 1_G = x$.
- ***Existence of inverses:*** *for all* $x \in G$*, there exists a unique* $x^{-1} \in G$ *such that* $x * x^{-1} = x^{-1} * x = 1_G$.

If in addition the operation $*$ *is commutative, i.e.* $x * y = y * x$ *for all* x, y*, then we say that* G *is an* ***abelian group****.*

Examples of abelian groups include the integers $(\mathbb{Z}, +)$ with the operation of addition, the nonzero real numbers with the usual multiplication $(\mathbb{R} \setminus \{0\}, \cdot)$ and p-adic groups, described below. Nonabelian groups include the **symmetric group** on a set[2] $\mathcal{A}$, $(S_{\mathcal{A}}, \circ)$, that consists of all bijections from the set $\mathcal{A}$ to itself, and the **general linear group** $\mathrm{GL}_d(\mathbb{R})$, which consists of all invertible real $d \times d$ matrices, with the operation of matrix multiplication.

[1] As usual, we use infix notation for the group operation, writing $x*y$ instead of $*(x, y)$ or $*xy$. Furthermore, if the group operation is deemed understood by context, we omit any specific symbol for it and just write xy instead of $x * y$.

[2] Usually, $\mathcal{A} = \{1, 2, \ldots, n\}$; in this case, we write S_n for the corresponding symmetric group.

Definition 1.2 *Let G be a group and $H \subseteq G$ any subset. We say that H is a* ***subgroup*** *of G (symbolized*[3] *as $H \leq G$) if H is itself a group with the restriction of $*$ to H as a operation, i.e. if it satisfies the following three properties:*

- ***Closure:*** *if $x, y \in H$, then $x * y \in H$.*
- ***Neutral element:*** *$1_G \in H$.*
- ***Closure of inverses:*** *for all $x \in H$, the element x^{-1} is also in H.*

If $H \leq G$, then for any $g \in G$ the following subset is called a ***left coset*** *of H:*

$$g * H = \{g * h \ : \ h \in H\}.$$

Right cosets $H * g$ are defined similarly. If for all $g \in G$ the equality $g * H = H * g$ holds, we say that H is a ***normal subgroup*** *of G, written $H \trianglelefteq G$. The set of all left (resp. right) cosets of a subgroup $H \leq G$ is written as G/H (resp., $H \backslash G$); when H is normal, both of these sets are equal.*

For finite groups, we usually use the term **order** for its cardinality $|G|$. If H is a subgroup of G, its **index** $[G : H]$ is the cardinality of the set G/H (or $H \backslash G$, as both sets have the same cardinality).

A very simple criterion to determine which subsets can be subgroups is the following:

Theorem 1.3 (Lagrange) *If H is a subgroup of G, then $|H|$ divides $|G|$.*

Corollary 1.4 *Let G be a finite group. For any $g \in G$, the equality $g^{|G|} = 1_G$ holds.*

The latest observation comes from the fact that $H = \{g^n \ : \ n \in \mathbb{Z}\}$ is a subgroup of G, implying that it is finite and that $g^n = g^{n+|H|}$ for all n. This $|H|$ is also called the **order** of the element $g \in G$ and often written as $\mathrm{ord}(g)$, as it is the least positive n such that $g^n = 1_G$.

1.2. Group homomorphisms

We often want to compare different groups. For instance, if $(\mathbb{R}^+, \cdot)$ is the group of positive real numbers under multiplication, we see that every positive real number $x \in \mathbb{R}^+$ can be written as e^t for some $t \in \mathbb{R}$, and this is a bijection between $\mathbb{R}$ and $\mathbb{R}^+$. Furthermore, using exponentials allows us to describe multiplication more easily via the usual exponent law $e^t \cdot e^s = e^{t+s}$, converting multiplication into addition, which is a different binary operation. Thus, in a way, multiplication and addition are "the same". To state this equivalence formally, we need to introduce operation-preserving functions as a definition:

Definition 1.5 *Let $(G, \cdot)$ and $(H, *)$ be two groups. A* ***group homomorphism*** *(or just* ***group morphism****) is a function $f \colon G \to H$ that satisfies the identity:*

$$(\forall g, h \in G) \colon f(g \cdot h) = f(g) * f(h).$$

[3] As usual, we use $H < G$ to symbolize "$H \leq G$ and $H \neq G$". We use $H \triangleleft G$ for strict normal subgroups as well. To keep consistency, we also use $H \subset G$ for strict set inclusion, instead of alternative notations such as $H \subsetneq G$.

If f is injective, we say it is a ***monomorphism*** *(often written as $f\colon G \hookrightarrow H$), and similarly we call an epijective homomorphism an* ***epimorphism*** *(and we use a two-headed arrow, $f\colon G \twoheadrightarrow H$). When f is a bijection, we call it an* ***isomorphism*** *and say $(G, \cdot)$ and $(H, *)$ are* ***isomorphic groups****; in this case, we write $(G, \cdot) \equiv (H, *)$.*

We think of isomorphic groups as different versions of the same "abstract group", since every property held by one of them is shared by the other. For instance, the group of the symmetries of a triangle, D_3, is isomorphic to the set of all permutations of the set $\{1, 2, 3\}$, S_3.

Composition of group homomorphisms is once again an homomorphism, and, in particular, it is easy to verify that the set of all isomorphisms of a group G to itself (that is, **automorphisms**) is a group, called the **automorphism group** of G and denoted $\mathrm{Aut}(G)$.

Group morphisms define some subgroups of interest both in their domain and codomain. We shall be mainly interested in the following:

Definition 1.6 *Let $f\colon G \to H$ be a group homomorphism. The* ***kernel*** *of f is the preimage under f of 1_H, that is:*

$$\ker(f) := f^{-1}[\{1_H\}]\{g \in G \;:\; f(g) = 1_H\}.$$

The ***image*** *of f is the set of all elements of H that are images of some $g \in G$ under f, i.e.*

$$\mathrm{im}(f) := f[G] = \{f(g) \;:\; g \in G\}.$$

The kernel $\ker(f)$ is always a normal subgroup of G, with the property that $f(g) = f(h)$ if and only if $gh^{-1} \in \ker(f)$. The image is a subgroup of H, although not always normal. The group homomorphism f is injective if and only if $\ker(f) = \{1_G\}$, and is surjective if and only if $\mathrm{im}(f) = H$.

1.3. Group actions

Usually groups are defined in terms of transformations of spaces or other objects, e.g. the set of all homeomorphisms from a topological space X to itself forms a group, $\mathrm{Homeo}(X)$, and the elements of this group have a very special relationship with the elements (points) of the space X. Similarly, the group of $d \times d$ invertible matrices over a field $\mathbb{K}$, $\mathrm{GL}_d(\mathbb{K})$, has an interesting relationship with the corresponding vector space $\mathbb{K}^d$, dictated by matrix multiplication; we briefly detail the main aspects of this connection between abstract groups and sets of transformations of spaces. The interested reader should consult the book by de Neymet [32].

Definition 1.7 *Let G be a group and X be a set. A* ***(left) group action*** *of G on X is a function $\varphi\colon G \times X \to X$ that satisfies the following properties:*

- *$\varphi(1_G, x) = x$, for every $x \in X$,*
- *$\varphi(g, \varphi(h, x)) = \varphi(gh, x)$, for every $g, h \in G, x \in X$.*

We write $G \overset{\varphi}{\curvearrowright} X$ to state that $\varphi\colon G \times X \to X$ is a group action of G on X.

We often write $g \cdot x$ instead of $\varphi(g, x)$, and we write $\varphi_g \colon X \to X$ for the function $\varphi_g(x) := \varphi(g, x)$. We note that, since $\varphi_{1_G} = \mathrm{id}_X$ and $\varphi_g \circ \varphi_{g^{-1}} = \varphi_{1_G} = \mathrm{id}_X$, every φ_g is a bijection $X \to X$ and thus an element of the symmetric group S_X. Thus, we may think of a group action $G \overset{\varphi}{\curvearrowright} X$ as a group morphism $\tilde{\varphi} \colon G \to S_X$. In particular, any subgroup of S_X acts naturally on the set X in an obvious way.

A group action $G \overset{\varphi}{\curvearrowright} X$ partitions the set X into "orbits", which are subsets $Y \subseteq X$ such that every point from Y is "reachable" from any other point of Y via the group action. More precisely:

Definition 1.8 *Let $G \overset{\varphi}{\curvearrowright} X$ be a group action. The **orbit** of a point $x \in X$ is the following set:*

$$\mathrm{Orb}(x) = G \cdot x := \{g \cdot x \ : \ g \in G\}.$$

Orbits are equivalence classes under the relation given by:

$$x \sim_\varphi y \iff (\exists g \in G) \colon g \cdot x = y,$$

and thus we write X/G for the set of all orbits. If $|X/G| = 1$ (i.e. for some, and thus all, $x \in X$, we have $\mathrm{Orb}(x) = X$*), we say that the group action is **transitive**.*

*A **fundamental domain** for the group action φ is a subset $D \subseteq X$ such that $|D \cap E| = 1$ for all $E \in X/G$, that is, it is a set of representatives of the equivalence classes for $\sim_\varphi$.*

Since the set X often has some structure, it is common to introduce weaker versions of the notion of transitivity. For example, if X is a topological space, we say that φ is **topologically transitive** if some orbit $\mathrm{Orb}(x)$ is dense in X. Note that now it might be the case that not every orbit is dense, and thus we also say that x is a **transitive point**; if every point is transitive, we say that the action φ is **minimal**.

We can also give stronger notions of transitivity. For instance, since a group action is transitive if, and only if, for every $x, y \in X$ there exists some $g \in G$ such that $g \cdot x = y$, we may instead check this property for several points simultaneously. Thus, a group action is **doubly transitive** if for every four points $x_1 \neq x_2, y_1 \neq y_2 \in X$ there exists some $g \in G$ such that $g \cdot x_i = y_i, i = 1, 2$. This easily generalizes to n pairs of points.

A notion closely related to that of orbit is the following subgroup of G:

Definition 1.9 *Let $G \overset{\varphi}{\curvearrowright} X$ be a group action and $x \in X$ any point. The **stabilizer** of x is the following subgroup of G:*

$$\mathrm{Stab}(x) = G_x := \{g \in G \ : \ g \cdot x = x\}.$$

The orbit and stabilizer relate to each other via the following well-known result:

Theorem 1.10 (Orbit-stabilizer theorem) *Let $G \overset{\varphi}{\curvearrowright} X$ be a group action and $x \in X$ be any point. Then, the following equality holds:*

$$|\mathrm{Orb}(x)| = [G : \mathrm{Stab}(X)].$$

This is a particularly strong result in the case where the set X is finite, or when all orbits are finite. It still holds true when X is infinite, in the form of the cardinal equality $|G| = |\mathrm{Orb}(x)| \cdot |\mathrm{Stab}(x)|$, although, due to how cardinal arithmetic works, the end result may appear "less interesting" in general.

Stabilizers appear often in connection with repetitive behavior and periodicity. A **period** of a point $x \in X$ is an element $g \in \mathrm{Stab}(x) \setminus \{1_G\}$. The following notions are closely related:

Definition 1.11 *A group action $G \overset{\varphi}{\curvearrowright} X$ is* ***faithful*** *if $\bigcap_{x \in X} \mathrm{Stab}(x) = \{1_G\}$, i.e. if for every $g \in G \setminus \{1_G\}$ there is some $x \in X$ such that $g \cdot x \neq x$. Similarly, φ is* ***free*** *if, for every $x \in X$ and $g \in G \setminus \{1_G\}$, we have $g \cdot x \neq x$.*

Free group actions are always faithful. The latter property may be interpreted in terms of the associated homomorphism $\tilde{\varphi} \colon G \to S_X$, which is injective if and only if the action is faithful. Alternatively, one may say that a group action is faitful if the points from X do not all share a common period.

As stated before, the set X may have some additional structure, and thus we may want to impose some constraints on the group action φ in order to ensure a good interplay between it and the structure of the set G. For instance, when X is a topological space, one would like all mappings φ_g to be continuous, and, since $\varphi_{g^{-1}} = \varphi_g^{-1}$, homeomorphisms of the space X into itself. Similarly, we may think of group actions on the vertices or edges of a graph; in such a case, we impose that adjacency must be preserved, i.e. if v_1, v_2 are two adjacent vertices (or edges), then $g \cdot v_1$ and $g \cdot v_2$ must be adjacent as well. More complex examples appear in several branches of mathematics.

One particularly interesting case is when X is itself a group; then, we request that every φ_g is a group homomorphism (isomorphism) $X \to X$, and say that G **acts on X by automorphisms**; the notation ${}^g x$ instead of $g \cdot x$ is common in this context. For instance, G acts on itself via **inner automorphisms**, via the group action defined by conjugation as follows:

$$ {}^g h = \varphi(g, h) := ghg^{-1}. $$

Any mapping of the form φ_g, in this case, is called an **inner automorphism** as well; the set of all inner automorphisms is a subgroup of $\mathrm{Aut}(G)$, often written $\mathrm{Inn}(G)$. Any automorphism of G not of this form is called an **outer automorphism**, and we write $\mathrm{Out}(G)$ for the quotient group $\mathrm{Aut}(G)/\mathrm{Inn}(G)$ (see the definition below).

1.4. Group constructions

There exist several constructions that create new groups from known ones. We dedicate this section to briefly list the ones we are interested in for what follows.

1.4.1. Intersections, unions and generating sets

Since subgroups are closed under the group operation, if g_1 and g_2 both belong to two different subgroups $H_1, H_2 \leq G$, then $g_1 g_2$ belongs to both subgroups as well, and is thus in their intersection. A similar reasoning for identity and inverse elements thus shows that $H_1 \cap H_2$ is also a subgroup of G. This immediately extends to arbitrary collections of subgroups:

Proposition 1.12 *Let $\{H_i\}_{i\in I}$ be a family of subgroups of G. Then $H = \bigcap_{i\in I} H_i$ is a subgroup of G as well. Furthermore, if all H_i are normal subgroups of G, then $H \trianglelefteq G$.*

Unions of subgroups usually do not have this property (much less unions of arbitrary groups); in fact, if $H_1 \cup H_2$ is a subgroup of G, then $H_1 \leq H_2$ or $H_2 \leq H_1$. However, in some circumstances the union of a specific kind of infinite family of subgroups is a new subgroup:

Proposition 1.13 *Let $\{H_i\}_{i=0}^{\infty}$ be a denumerably infinite family of groups that form an* ***ascending chain****, i.e. $H_0 \leq H_1 \leq H_2 \leq \dots$. Then $H = \bigcup_{i\in I} H_i$ is a group. If all the H_i are subgroups of a given group G, then $H \leq G$.*

Below, we shall mention other group constructions that generalize directed unions, and that end up being very important for our purposes.

The fact that arbitrary intersections of subgroups are themselves subgroups allows us to give answer to questions such as "which is the smallest subgroup (under inclusion) of G that contains a given element $g \in G$?" The general definition is as follows:

Definition 1.14 *Let G be a group and $S \subseteq G$ be any subset of G. The* ***subgroup generated by*** *S is the smallest subgroup under inclusion that contains S, which equals:*

$$\langle S\rangle := \bigcap_{S\subseteq H\leq G} H.$$

Similarly, the ***normal subgroup generated by*** *S is the smallest normal subgroup that contains S, equal to:*

$$\langle\langle S\rangle\rangle := \bigcap_{S\subseteq H\trianglelefteq G} H.$$

If $G = \langle S\rangle$ for some $S \subseteq G$, we say that S is a ***generating set*** *for G, and we say that S is* ***irredundant*** *if $\langle T\rangle \neq \langle S\rangle$ for any strict subset $T \subset S$. The least cardinality of a generating set for G is called the* ***rank*** *of the group G and written* $\mathrm{rank}(G)$.

If $G = \langle S\rangle$, then every element of G may be written as a finite product of elements of S and their inverses. In particular, a group of rank 1 (i.e. generated by a single element c) is called a **cyclic group**, and is always isomorphic to $\mathbb{Z}$ or to $\mathbb{Z}/n\mathbb{Z}$ for some $n \geq 1$, as all of its elements are of the form c^k for some $k \in \mathbb{Z}$; thus, they are either finite or denumerably infinite. Similarly, a **finitely generated group** (i.e. of finite rank) is always denumerable.

Group homomorphisms are strongly related to generating sets: as every element $g \in G$ is a product of $s_i^{\pm 1}$ with the $s_i \in S$, the element $f(g)$ is a product of the corresponding $f(s_i)^{\pm 1}$. Thus, we have the following property:

Proposition 1.15 *Let $f, g\colon G \to H$ be two group homomorphisms and suppose that $G = \langle S\rangle$. If $f|_S = g|_S$, then*[4] *$f = g$.*

[4] However, it is not always the case that a map $S \to H$ extends to a group morphism $G \to H$. A **free group** G is one with this property, i.e. G has a generating set S such that every map $S \to H$ extends to a group morphism $G \to H$; for example, $\mathbb{Z}$ is a free group of rank 1, but any other cyclic group is not free. We shall not deal with free groups in any big capacity in the rest of this work.

1.4.2. Quotient groups

If H is a normal subgroup of G, we may give a group structure to the set of cosets $G/H = H\backslash G$, as follows:

Definition 1.16 *Let H be a normal subgroup of G, and $g, h \in G$. We say that g is* ***congruent*** *to h modulo H if $gh^{-1} \in H$, and write $g \equiv h \pmod{H}$; this is an equivalence relation, whose equivalence classes are exactly the cosets gH from the set G/H. It is not hard to verify that if $g \equiv g', h \equiv h' \pmod{H}$, then $gh \equiv g'h' \pmod{H}$, and thus this defines a binary operation on G/H given by:*

$$gH * hH = (gh)H.$$

This operation satisfies the properties of a group listed above, and thus we call G/H jointly with this binary operation the ***quotient group*** *given by the subgroup H.*

There is a natural group epimorphism $p\colon G \twoheadrightarrow G/H$ that maps every $g \in G$ to its corresponding equivalence class under congruence (that is, the coset gH). This quotient map actually gives a strong characterization of quotient groups, as follows:

Theorem 1.17 (Factor theorem, or fundamental homomorphism theorem) *Let $f\colon G \to H$ be a group morphism and $N \trianglelefteq G$ a normal subgroup. Suppose that $N \subseteq \ker(f)$, that is, $f(g) = 1_H$ for every $g \in N$. Then there exists a map $\tilde{f}\colon G/N \to H$ such that $f = \tilde{f} \circ p$, i.e. the following diagram commutes:*

$$\begin{tikzcd}
G \arrow[r, "f"] \arrow[d, "p"'] & H \\
G/N \arrow[ur, dashed, "\tilde{f}"'] &
\end{tikzcd}$$

It is easy to see that $\ker(\tilde{f}) = \ker(f)/N$. Thus, this immediately leads to the following well-known result:

Corollary 1.18 (First isomorphism theorem) *Given a morphism $f\colon G \to H$, there is a canonical isomorphism:*

$$G/\ker(f) \cong \operatorname{im}(f),$$

which is given by the map $\tilde{f}$ from the factor theorem.

1.4.3. Direct and semidirect products, wreath products and extensions

Remember that the **Cartesian product** of two sets A and B is the set of all ordered pairs (a, b) where $a \in A$ and $b \in B$. This construction generalizes to any collection of sets, even infinite ones[5]: if $\{A_i\}_{i\in I}$ is some collection of sets, $\prod_{i\in I} A_i$ is the set of all functions[6]

[5] In the infinite case, one may need to appeal to the Axiom of Choice to ensure that the Cartesian product of infinitely many nonempty sets is nonempty; however, we do not need the full generality of this construction in the cases we shall deal with below.

[6] We use the notation a_i instead of $a(i)$ for the "i-th coordinate" of an element of a Cartesian product.

$a\colon I \to \bigcup_{i\in I} A_i$ such that $a_i \in A_i$ for all $i \in I$. It is easy to see that an ordered tuple $(a_1, \ldots, a_n)$ may be seen as a function $\{1, \ldots, n\} \to A_1 \cup \cdots \cup A_n$ that satisfies the condition $a_i \in A_i$, which shows how this definition is a generalized version of finite Cartesian products.

When the individual sets involved in a Cartesian product are groups, there is an obvious way to give a group structure to the corresponding product, which corresponds to "operating coordinate-wise", as follows:

Definition 1.19 *Given a family of groups $\{G_i\}_{i\in I}$, the **(external)**[7] **direct product** of this family is the set $G = \prod_{i\in I} G_i$, together with the binary operation $*$ given by:*

$$(\forall g, h \in G, i \in I)\colon (g * h)_i = g_i h_i.$$

*Relatedly, the **direct sum** of the family $\{G_i\}_{i\in I}$ is the following subgroup of $\prod_{i\in I} G_i$:*

$$\bigoplus_{i\in I} G_i := \left\{ (g_i)_{i\in I} \in \prod_{i\in I} G_i \ : \ (\exists F \subseteq I \text{ finite})(\forall i \in I \setminus F)\colon g_i = 1_{G_i} \right\},$$

that is, the elements of $\bigoplus_{i\in I} G_i$ are those elements of $\prod_{i\in I} G_i$ that are the identity in all but a finite number of coordinates. When I is finite, both groups are the same.

In a direct product of two groups, the subgroups $\tilde{G} = G \times \{1_H\}$ and $\tilde{H} = \{1_G\} \times H$ are copies of G and H, normal in $G \times H$, and every element of $G \times H$ is of the form $\tilde{g}\tilde{h}$, with $\tilde{g} \in \tilde{G}, \tilde{h} \in \tilde{H}$. Thus, we might say that $G \times H$ is a group that "decomposes" into G and H, and the latter two groups describe the whole structure of $G \times H$. The purpose of most of these constructions is to express the structure of a new group we may encounter in terms of groups we already know, as in this example, or in the case of quotient groups.

However, any element from $\tilde{G}$ commutes with any other element from $\tilde{H}$. For our "deconstruction" process, this restriction might be excessive in some circumstances. A situation that often appears is that the estructure of some group is given entirely by two subgroups G and H, but only one is normal in the larger group. We give this construction a name, as follows:

Definition 1.20 *Let $(G, *)$ and $(H, \star)$ be two groups, and suppose there is a group action $H \overset{\varphi}{\curvearrowright} G$ by automorphisms. Define the following binary operation on the set $G \times H$:*

$$(g_1, h_1) \cdot (g_2, h_2) = (g_1 * {}^{h_1}g_2, h_1 \star h_2).$$

The set $G \times H$, endowed with this binary operation, forms a group, which is called the (external)[8] ***semidirect product** of G and H. We write $G \rtimes_\varphi H$ for the resulting group structure, to distinguish it from the usual Cartesian product.*

It is not hard to see that once again $\tilde{G} = G \times \{1_H\}$ and $\tilde{H} = \{1_G\} \times H$ are both subgroups of $G \rtimes_\varphi H$, respectively isomorphic to G and H, and the first is normal in $G \rtimes_\varphi H$, while

[7] If a group G has a family of normal subgroups $\{G_i\}_{i\in I}$ with $G_i \cap G_j = \{1_G\}$ for all $i \neq j$ and $\langle \bigcup_{i\in I} G_i \rangle = G$, then we say that G is the **internal direct product** of the G_i; in this scenario, G is isomorphic to $\prod_{i\in I} G_i$ as defined above.

[8] Once again, if G has two subgroups $G_1 \trianglelefteq G, G_2 \leq G$ with $G_1 \cap G_2 = \{1_G\}$ and $\langle G_1 \cup G_2 \rangle = G$, the group G is isomorphic to the semidirect product $G_1 \rtimes_\varphi G_2$, where the action $G_2 \overset{\varphi}{\curvearrowright} G_1$ is given by $\varphi_g(x) = gxg^{-1}$.

the second is not unless the group action φ is trivial (in which case the semidirect product becomes just a direct product); thus, in a similar fashion to the direct product, elements of $G \rtimes_{\varphi} H$ decompose as products $\tilde{g}\tilde{h}$ with $\tilde{g} \in \tilde{G}, \tilde{h} \in \tilde{H}$, and now elements of $\tilde{G}$ no longer necessarily commute with those of $\tilde{H}$, showing the larger generality of this construction.

One particular case that is often encountered is the one in which G is a direct product (or direct sum) of copies of the same group, indexed by some set I, and H acts on the set of indices in any way, which induces a group action by automorphisms on G. More precisely:

Definition 1.21 *Let $H \overset{\varphi}{\curvearrowright} I$ be any group action of a group H on an arbitrary set I, and let G be any other group. The (unrestricted)*[9] ***wreath product*** *of G and H, written $G \wr H$, is the semidirect product $G^I \rtimes_{\hat{\varphi}} H$ (where $G^I = \prod_{i \in I} G$) where the action by automorphisms $H \overset{\hat{\varphi}}{\curvearrowright} G^I$ is given by:*

$$\hat{\varphi}(h, (g_i)_{i \in I}) = (g_{\varphi(h,i)})_{i \in I},$$

that is, the group operation is given by:

$$((g_i)_{i \in I}, h) \cdot ((g'_i)_{i \in I}, h') = ((g_i g'_{\varphi(h,i)}), hh').$$

One particular (and very important) case of a wreath product is the set of rigid symmetries of a d-dimensional cube W_d, which is isomorphic to the set of all graph automorphisms of the graph with vertex set $\{0,1\}^d$ and an edge joining v_1 and v_2 if and only if they differ in exactly one coordinate. This group can be seen to be isomorphic to $(\mathbb{Z}/2\mathbb{Z}) \wr S_d$; the informal reasoning for this equality is that any rigid symmetry of the cube has to permute the d coordinate axes in some way (corresponding to the permutation group component S_d), and some of them might be reflected afterwards, which corresponds to a sign change in some coordinate (and thus to the $(\mathbb{Z}/2\mathbb{Z})^d$ component, interpreting each $\mathbb{Z}/2\mathbb{Z}$ coordinate as the multiplicative group $\{+1, -1\}$). A similar reasoning shows that W_d is isomorphic to the group of all $d \times d$ signed permutation matrices.

Semidirect products (and thus, in particular, wreath products) are particular cases of the following general situation:

Definition 1.22 *Let G be any group, with $H \trianglelefteq G$ a normal subgroup and F any arbitrary group. We say that G is a H-by-F* ***group extension*** *if $G/H \cong F$, that is, if they fit in the following* ***short exact sequence****:*

$$\begin{tikzcd} \{1\} \arrow[r] & H \arrow[r, hook, "\iota"] & G \arrow[r, two heads, "p"] & F \arrow[r] & \{1\} \end{tikzcd}, \tag{1.1}$$

where ι is the inclusion map $H \hookrightarrow G$ and p is the quotient map $G \to G/H \cong F$; the term ***exact sequence*** *refers to the fact that the kernel of every group homomorphism in the sequence equals the image of the previous homomorphism in the sequence (thus, ι is injective, p must be surjective and $\ker(p) = \operatorname{im}(\iota)$).*

We say that this sequence is an ***split exact sequence*** *(or that it splits) if there exists a right inverse $\tau \colon F \to G$ for p, that is, a group morphism such that $p \circ \tau = 1_H$.*

[9] The same construction, using the direct sum instead of the direct product of copies of G, yields a specific subgroup of $G \wr H$ known as the **restricted wreath product**. Since we shall only encounter wreath products with finite index sets I, we do not need to deal with this distinction.

The importance of split exact sequences lies in the following theorem:

Theorem 1.23 *Let the group G be an H-by-F extension. If the associated short exact sequence splits, then G is isomorphic to a semidirect product of the form $H \rtimes_\varphi F$. The converse is also true.*

This result comes from the fact that $\tilde{F} = \mathrm{im}(\tau)$ is a copy of F contained in G, and thus it can be made to act on H by simple conjugation. The exactness of the associated sequence shows that $\tilde{F} \cap H = \{1_G\}$; the end result follows from there. Thus, a way to show that a given group is a semidirect product of previously known ones is to show that it is an extension, and then show either τ or the associated subgroup $\mathrm{im}(\tau)$; we shall make use of this technique in later chapters.

An example of a group extension that is not isomorphic to a semidirect product is $\mathbb{Z}$, as it may be seen as a $\mathbb{Z}$-by-$(\mathbb{Z}/2\mathbb{Z})$ extension, taking $H = 2\mathbb{Z} \cong \mathbb{Z}$; were $\mathbb{Z}$ isomorphic to such a direct product, it would contain elements of order 2, which is absurd.

1.4.4. Projective limits

The final example of a group construction we shall deal with is a very versatile one, of which previously detailed ones may be seen as subcases. As this is a very specialized construction, the reader may be inclined to consult a standard book on category theory [74,86] for a more in-depth description; we also recommend to consult de Neymet's book on group actions [32] for applications of this construction in the construction of dynamics-related groups. Remember that a **directed set** is a set I together with a binary relation $\preceq$ which is:

- **reflexive**, i.e. $x \preceq x$ for every $x \in I$,
- **transitive**, that is $x \preceq y \wedge y \preceq z \implies x \preceq z$, and
- **concurrent**, that is, for every $x, y \in I$ there exists some $z \in I$ such that $x \preceq z$ and $y \preceq z$.

Let $\{G_i\}_{i\in I}$ be any family of groups, where I is a directed set under the relation $\preceq$. We say that this is an **inverse system** of groups if there exists a family of group morphisms $f_{i,j}\colon G_j \to G_i$, defined for every pair of indices such that $i \preceq j$, that satisfies the following properties:

- $f_{i,i} = \mathrm{id}_{G_i}$, and
- $f_{i,j} \circ f_{j,k} = f_{i,k}$ whenever $i \preceq j \preceq k$.

Definition 1.24 *The **projective limit** (or **inverse limit**) of the inverse system of groups $(\{G_i\}_{i\in I}, \{f_{i,j}\}_{i\preceq j})$ is the following subgroup of $\prod_{i\in I} G_i$:*

$$\varprojlim_i G_i := \left\{ (g_i)_{i\in I} \in \prod_{i\in I} G_i \ : \ (\forall i \preceq j \in I)\colon f_{i,j}(g_j) = g_i \right\}.$$

If $G = \varprojlim_i G_i$, one may define the obvious proyection homomorphism $\pi_i\colon G \to G_i$ which satisfies the relation $\pi_j((g_i)_{i\in I}) = g_j$. We see that, by the condition we set on the elements of G, these projections satisfy the equality $\pi_i = f_{i,j} \circ \pi_j$. This property actually describes[10] projective limits entirely; that is, if H is another group that has morphisms $\beta_i \colon H \to G_i$ which satisfy the condition $\beta_i = f_{i,j} \circ \beta_j$ for every $i \preceq j$, then there is a morphism $\beta \colon H \to G$ such that $\beta_i = \pi_i \circ \beta$.

Projective limits can be seen as a generalization of infinite decreasing intersections of groups. Indeed, $\mathbb{N}$ with the usual ordering, $\leq$, may be seen as a directed set, and if $G_0 \geq G_1 \geq G_2 \geq \dots$ is a descending chain of subgroups of a group G, we may define $f_{i,j} \colon G_j \to G_i$ as the usual inclusion map whenever $i \leq j$; evidently, $f_{i,j} \circ f_{j,k} = f_{i,k}$. If we examine the inverse limit, we see that the following equality holds:

$$\varprojlim_{i} G_i = \left\{ (g_i)_{i \in \mathbb{N}} \in \prod_{i \in I} G_i \ : \ (\forall i \leq j \in \mathbb{N}) \colon g_i = g_j \right\},$$

since every G_i can be identified with a subgroup of all preceding $G_j, j < i$, and the inclusion maps send every element of G_i to itself. This means that every element of the inverse limit is an infinite sequence of equal elements $(g, g, g, \dots)$, and such a g has to belong to every G_i; reciprocally, if $g \in \bigcap_{i \in \mathbb{N}} G_i$, the infinite sequence $(g, g, g, \dots)$ satisfies the condition in the definition of the inverse limit. Thus, there is a natural bijection:

$$\begin{aligned} \varphi \colon \bigcap_{i \in \mathbb{N}} G_i &\to \varprojlim_{i} G_i, \\ g &\mapsto (g, g, g, \dots), \end{aligned}$$

and it is easy to see that this preserves the group operation and is thus an isomorphism.

One may allow distinct configurations of groups and homomorphisms and define projective limits in full generality without involving directed sets in the construction. For instance, it can be shown that both the direct and semidirect products may be seen as specific subcases of a more general definition of projective limit. However, we shall only need this restricted definition for our purposes, so we won't go into further detail.

1.5. Center, centralizer and normalizer

In this work we shall be interested in certain subgroups of a given group G which have an interesting algebraic description. Thus, we briefly describe them and their main properties.

Definition 1.25 *Let G be any group. The **center** of G is the subgroup $Z(G)$ of all elements $g \in G$ that commute with every other element of the group, i.e.:*

$$Z(G) := \{ g \in G \ : \ (\forall h \in G) \colon gh = hg \}.$$

$Z(G)$ is always an Abelian subgroup of G. Note that, if $z \in Z(G)$, then $zgz^{-1} = zz^{-1}g = g$ for all $g \in G$, and thus the inner automorphism φ_z defined previously is trivial. Reciprocally, if $\varphi_z(g) \neq g$, we have that $zgz^{-1} \neq g \implies zg \neq gz$, and thus $z \notin Z(G)$. This is the essence of the proof of the following:

[10] This is called a **universal property**.

Proposition 1.26 *For any group G, $\mathrm{Inn}(G) \cong G/Z(G)$. In particular, Abelian groups do not have nontrivial inner automorphisms.*

Closely related is the question of, given a specific subset $S \subseteq G$, which is the largest subgroup $H \leq G$ which commutes with every $s \in S$. We define:

Definition 1.27 *Let S be any subset*[11] *of G. The **centralizer** of S is the set of all elements of G which commute with every element of S:*

$$\mathrm{cent}(S) = \mathrm{cent}_G(S) := \{g \in G \ : \ (\forall s \in S) \colon gs = sg\}.$$

Note that $\mathrm{cent}_G(S)$ is actually a group, and that it may not actually contain elements from S; for instance, if $gh \neq hg$, $\mathrm{cent}_G(\{g, h\})$ is nonempty (as it at least contains 1_G) but cannot contain g (as it does not commute with h) nor h (as it does not commute with g). However, it is not hard to see that $\mathrm{cent}_G(\mathrm{cent}_G(S))$ does in fact contain S.

Note, finally, that $Z(G)$ is the largest subgroup $H \leq G$ for which $\mathrm{cent}_G(H) = G$.

Proposition 1.28 *For any $S \subseteq G$, we have $\mathrm{cent}_G(S) = \mathrm{cent}_G(\langle S \rangle)$.*

Proposition 1.29 *For any family of subsets $\{S_i\}_{i \in I}$ of G, we have:*

$$\mathrm{cent}_G\left(\bigcup_{i \in I} S_i\right) = \bigcap_{i \in I} \mathrm{cent}_G(S_i).$$

In particular, if H is a subgroup of G generated by $S = \{s_1, \ldots, s_k\}$, we can compute $\mathrm{cent}_G(H)$ as the intersection $\bigcap_{i=1}^{k} \mathrm{cent}_G(\{s_i\})$; for certain groups G, such as permutation groups, computing the centralizer of a single element can be done algorithmically, providing a reasonable method to compute centralizers of finitely generated subgroups.

Closely related to centralizers are the following:

Definition 1.30 *Let S be any subset of G. The **normalizer** of G is the following subset:*

$$\mathrm{norm}_G(S) = \{g \in G \ : \ gS = Sg\}.$$

In particular, if $H \leq G$ is a subgroup, $\mathrm{norm}_G(H)$ is the largest subgroup $H \trianglelefteq \mathrm{norm}_G(H) \leq G$ in where H is normal. We note that elements of $\mathrm{norm}_G(S)$ satisfy a sort of “quasi-commutativity”, in the sense that if $g \in G, s \in S$, there exists some $\tilde{s} \in S$ such that $gs = \tilde{s}g$. In particular, if H is a subgroup of G, there is an automorphism $f_g \in \mathrm{Aut}(H)$ such that $gh = f_g(h)g$ for every $g \in G, h \in H$, where f_g depends only on g; this fact will be important in later chapters.

[11] Often S is assumed to be a subgroup; however, it is convenient to allow any arbitrary set (in particular, singletons) in this definition. Replacing S by $\langle S \rangle$ results in the same set, so the chosen convention makes no real difference.

1.6. Some groups that will be encountered later

1.6.1. Permutation groups

As stated above, the **permutation group** associated to a set X is the group S_X of all bijections $X \to X$, with the operation of composition; these bijections are usually referred to as **permutations**. Often (but not always) X is taken to be the set $\{1, 2, \ldots, n\}$, and we use the notation S_n instead of S_X in such a situation; these groups are thoroughly described in the textbook by Hall [49]. Note that, in this case, $|S_n| = n!$.

As usual in group theory, we forgo the composition symbol whenever it does not lead to confusion. Thus, we will write $\sigma\tau$ instead of $\sigma \circ \tau$. There are two standard notations that are used for permutations on finite sets. The first one is pretty straightforward:

$$\sigma = \begin{pmatrix} 1 & 2 & \ldots & n \\ s_1 & s_2 & \ldots & s_n \end{pmatrix},$$

which means that σ is the bijection that maps the element i to the corresponding s_i. This notation lends itself nicely to direct computation of compositions and inverses of permutations; for instance:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}, \quad \tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix} \implies \sigma\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ {\scriptstyle 4} & {\scriptstyle 1} & {\scriptstyle 2} & {\scriptstyle 3} \\ 1 & 2 & 4 & 3 \end{pmatrix},$$

and similarly:

$$\sigma^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix}.$$

In both cases, computation is performed via simple operations such as matching columns from both permutations that share an entry, flipping or reordering the corresponding columns. However, this notation quickly becomes cumbersome for values of n larger than, say, 10, and thus it is often impractical. The second notation relies on the following definitions:

Definition 1.31 *Let $\sigma \in S_X$ be any permutation. The **support** of σ is the following subset of X:*

$$\mathrm{supp}(\sigma) = \{x \in X \ : \ \sigma(x) \neq x\},$$

that is, the complement of $\mathrm{supp}(\sigma)$ *is the largest set where σ acts trivially.*

For instance, the support of the permutation σ of the previous example is $\mathrm{supp}(\sigma) = \{1, 2, 4\}$, since $\sigma(3) = 3$. Note that σ necessarily maps $\mathrm{supp}(\sigma)$ to itself.

Definition 1.32 *A **cycle** (of length k) is a permutation $\sigma \in S_X$ with finite support* $\mathrm{supp}(\sigma) = \{s_1, \ldots, s_k\}$ *such that $\sigma(s_i) = s_{i+1}$ for all $1 \leq i \leq k-1$, and $\sigma(s_k) = s_1$; a **transposition** is a cycle of length 2. Two cycles $\sigma, \tau \in S_X$ are called **disjoint** if* $\mathrm{supp}(\sigma) \cap \mathrm{supp}(\tau) = \varnothing$.

The standard notation for a cycle σ as described above is:

$$\sigma = (s_1 \, s_2 \, \ldots \, s_k).$$

Note that $(s_j \, s_{j+1} \, \ldots \, s_k \, s_1 \, s_2 \, \ldots \, s_{j-1})$ represents the same cycle, for any value of j between 1 and k.

Lemma 1.33 *If σ and τ are disjoint cycles, then $\sigma\tau = \tau\sigma$.*

This comes as a consequence of the equality $\sigma[\mathrm{supp}(\sigma)] = \mathrm{supp}(\sigma)$. Thus, τ behaves like the identity in $\mathrm{supp}(\sigma)$, implying $\tau(\sigma(x)) = \sigma(x) = \sigma(\tau(x))$ for all $x \in \mathrm{supp}(\sigma)$; the same reasoning holds in reverse by swapping the roles of σ and τ. Everywhere else, $\sigma\tau(x) = x = \tau\sigma(x)$ trivially.

Theorem 1.34 (Cycle decomposition theorem) *If $|X| < \infty$, every permutation is a composition of disjoint cycles in a unique way up to ordering.*

To construct the individual cycles, we choose any $x \in \mathrm{supp}(\sigma)$ and define $x_0 = x, x_1 = \sigma(x), x_2 = \sigma^2(x)$, and so on. Since X is finite and σ is a bijection, eventually $x_k = x_0 = x$ for some $k > 0$, and thus the restriction of σ to $\{x_1, \dots, x_n\}$ is equal to the cycle $\kappa_1 = (x_1\, x_2\, \dots\, x_n)$. Iterating this process with some $y \in \mathrm{supp}(\sigma) \setminus \mathrm{supp}(\kappa_1)$ and so on produces the desired cycles $\kappa_2, \kappa_3, \dots,$ which have disjoint support and commute pairwise.

Thus, arbitrary permutations are often described with cycle notation using the above cycle decomposition. For instance:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 2 & 3 & 1 \end{pmatrix} = (1\,5)(2\,4\,3).$$

An important consequence of the cycle decomposition theorem is the following:

Corollary 1.35 *Every permutation with finite support is a finite product of transpositions. If σ can be written in two different ways, as a product of m transpositions and as a product of k transpositions, then m and k have the same parity, and we say that σ is an* ***even permutation*** *or* ***odd permutation*** *depending on whether m (and thus k) is even or odd.*

This is a consequence of the following equality:

$$(s_1\, s_2\, \dots\, s_k) = (s_1\, s_k)(s_1\, s_{k-1}) \cdots (s_1\, s_3)(s_1\, s_2),$$

and thus a cycle of length k is even if and only if k is odd.

Corollary 1.36 *S_n is generated by the set of all transpositions.*

Furthermore, we see that the product of two even permutations or two odd permutation is even, while the product of an even permutation with an odd permutation is an odd permutation. This motivates the following:

Definition 1.37 *The* ***alternating group*** *$A_n \leq S_n$ is the subset (subgroup) of all even permutations.*

We can easily see that A_n is effectively a group after noting that it is the kernel of the following group homomorphism $\mathrm{sgn}\colon S_n \to \{+1, -1\} \cong \mathbb{Z}/2\mathbb{Z}$:

$$f(\sigma) = \begin{cases} 0 & \text{if } \sigma \text{ is even,} \\ 1 & \text{if } \sigma \text{ is odd.} \end{cases}$$

Since the group $\{+1,-1\}$ has only two elements, $[S_n : A_n] = 2$. It can be proven that A_n can be generated by the set of all cycles of length 3. Furthermore, the following is a very important property of the alternating group:

Theorem 1.38 *For all values of n except $1, 2$ and 4, the group A_n is **simple**, i.e. the only normal subgroups it has are A_n itself and the trivial subgroup $\{1_{S_n}\}$.*

The following computational property is very useful when dealing with conjugation in permutation groups:

Proposition 1.39 *Let σ be any permutation and $\kappa = (s_1\, s_2\, \ldots\, s_k)$ be a cycle. We have that $\sigma\kappa\sigma^{-1}$ is a cycle as well; more precisely:*

$$\sigma(s_1\, s_2\, \ldots\, s_k)\sigma^{-1} = (\sigma(s_1)\, \sigma(s_2)\, \ldots\, \sigma(s_k)).$$

This can be used as part of an algorithm to compute centralizer and normalizers, whose exposition will be delayed until a later chapter. However, we note this interesting consequence:

Corollary 1.40 *For all $n \geq 3$, $\operatorname{rank}(S_n) = 2$. Indeed, $S_n = \langle (1\,2), (1\,2\,3\, \ldots\, n) \rangle$.*

Indeed, conjugating $(1\,2)$ by powers of $(1\,2\,3\, \ldots\, n)$ results in all transpositions of the form $(k\,(k+1))$. Conjugating $(2\,3)$ by $(1\,2)$ produces $(1\,3)$, which by conjugation again produces all transpositions $(k\,(k+2))$; iterating this process results in all permutations, which generate the whole of S_n.

Our interest in symmetry groups does not come only fr om their natural group action on a given set, but also from the following well-known result:

Theorem 1.41 (Cayley representation theorem) *Let G be any finite group and $n = |G|$. Then G is isomorphic to a subgroup of S_n, which consists of the bijections $L_g\colon G \to G$ (after identifying G with $\{1, 2, \ldots, n\}$ in some fashion) given by:*

$$L_g(x) = g \cdot x,$$

*for any $g \in G$. The subgroup $L(G)\{L_g : g \in G\}$ is called the **left Cayley representation** of G. Similarly, the functions:*

$$R_g(x) = x \cdot g^{-1},$$

*generate another isomorphic copy of G into S_n, called the **right Cayley representation** of G and written $R(G)$. Usually, $L(G) \neq R(G)$.*

It is obvious that $L_g \neq L_h$ if $g \neq h$, as $L_g(1_G) = g$. Besides, these functions satisfy the natural properties $L_g \circ L_h = L_{gh}$ and $L_g^{-1} = L_{g^{-1}}$, showing that the set of all L_g is a subgroup of S_n, after identifying the elements of G with $\{1, 2, \ldots, n\}$. Verifying the corresponding properties for $R(G)$ is straightforward.

An interesting property that will be useful later is the following:

Theorem 1.42 *For any group G with $n < \infty$ elements, its left and right Cayley representations are linked by the following relationship:*

$$R(G) = \mathrm{cent}_{S_n}(L(G)), \quad L(G) = \mathrm{cent}_{S_n}(R(G)).$$

This comes from the observation that $L_g(R_h(x)) = g(xh^{-1}) = (gx)h^{-1} = R_h(L_g(x))$. A similar, but slightly more complicated relationsip, holds for normalizers:

Theorem 1.43 *If G is any group of $n < \infty$ elements, $\mathrm{Aut}(G)$ is isomorphic to some subgroup $A \leq S_n$, after identifying G with $\{1, 2, \dots, n\}$, and A has trivial intersection with both $L(G)$ and $R(G)$. Furthermore, if $\sigma \in S_n$ is a permutation such that, for every $g \in G$, there exists some $h \in G$ such that $\sigma L_g \sigma^{-1} = L_h$, then there exist $\sigma' \in A$ (i.e. the representation of some automorphism of G) and $k \in G$ such that $\sigma = L_k \sigma'$. The converse also holds. Thus:*

$$\mathrm{norm}_{S_n}(L(G)) = \langle L(G) \cup A \rangle \cong G \rtimes \mathrm{Aut}(G) \cong \langle R(G) \cup A \rangle = \mathrm{norm}_{S_n}(R(G)).$$

1.6.2. Matrix groups

We assume the elementary properties of matrices, matrix operations (addition, multiplication, inversion, determinant, etc.) and so on to be known by the reader. We shall write $\mathbb{M}_d(\mathbb{K})$ for the set (which is actually a ring) of all $d \times d$ matrices with entries over a ring[12] $\mathbb{K}$. We also write I_d for the corresponding identity matrix. We are interested in subsets of $\mathbb{M}_d(\mathbb{K})$ closed under matrix multiplication and inversion, that is, that conform groups under matrix multiplication.

Definition 1.44 *Let $\mathbb{K}$ be a ring and $d \geq 1$. The **general linear group** $\mathrm{GL}_d(\mathbb{K})$ is the set of all $d \times d$ invertible matrices in $\mathbb{K}$. A **linear group** is one that is isomorphic to some subgroup of $\mathrm{GL}_d(\mathbb{K})$ for some d and $\mathbb{K}$.*

Remember that $\mathbb{K}^\times$ is the set of all **units** of $\mathbb{K}$ (i.e. elements which have inverses); in particular, when $\mathbb{K}$ is a field, $\mathbb{K}^\times \cong \mathbb{K} \setminus \{0\}$. We have the following characterization of the elements of the general linear group:

Theorem 1.45 *A matrix $A \in \mathbb{M}_d(\mathbb{K})$ belongs to $\mathrm{GL}_d(\mathbb{K})$ if, and only if, $\det(A) \in \mathbb{K}^\times$.*

For instance, $\mathrm{GL}_d(\mathbb{R})$ consists of all $d \times d$ matrices whose determinant is nonzero (equivalently, with linearly independent columns), while $\mathrm{GL}_d(\mathbb{Z})$ is the subgroup of $\mathrm{GL}_d(\mathbb{Z})$ that consists of matrices with integer entries and whose determinants are ± 1 (thus, not every matrix from $\mathrm{GL}_d(\mathbb{R})$ with integer coefficients belongs to $\mathrm{GL}_d(\mathbb{Z})$).

Some linear groups that appear fairly often include:

- the **special linear group**:

$$\mathrm{SL}_d(\mathbb{K}) := \{A \in \mathrm{GL}_d(\mathbb{Z}) \ : \ \det(A) = 1\},$$

[12] Which is often, but not always, expected to be a field. For the purposes of this work, we shall only be interested in rings of characteristic $\neq 2$, so we will not make note of the differences in treatment regarding, e.g., determinants.

- the **orthogonal group** (for real matrices) and the **unitary group** (its complex equivalent):

$$\mathrm{O}_d(\mathbb{R}) := \{A \in \mathrm{GL}_d(\mathbb{R}) \ : \ A^{\mathsf{T}}A = I_d\},$$
$$\mathrm{U}_d(\mathbb{C}) := \{A \in \mathrm{GL}_d(\mathbb{C}) \ : \ \overline{A}^{\mathsf{T}}A = I_d\},$$

- the **upper triangular group** (and its lower triangular counterpart):

$$\Delta_d(\mathbb{K}) := \{A \in \mathrm{GL}_d(\mathbb{K}) \ : \ A_{i,j} = 0 \text{ for all } i > j\},$$

- the **Heisenberg group**:

$$H(\mathbb{K}) = \left\{ \begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix} \ : \ a, b, c \in \mathbb{K} \right\},$$

  where $\mathbb{K}$ is often chosen as either $\mathbb{R}$ or $\mathbb{Z}$.

Part of our interest in matrix groups comes from the following notion:

Definition 1.46 *Let G be any group. A* ***linear representation*** *(of dimension d over $\mathbb{K}$) of G is any group homomorphism $\varrho \colon G \to \mathrm{GL}_d(\mathbb{K})$ for some d and $\mathbb{K}$; we say ϱ is* ***faithful*** *if it is injective.*

We won't go into detail regarding representation theory; however, it is important to note that a group G is linear if, and only if, it has a faithful representation. In a further chapter we shall be interested in linear representations of general rings $\mathbb{K}$, which are connected with the associated representations of their unit groups $\mathbb{K}^{\times}$; linear representation of groups and rings facilitates computation and might make certain features of the groups and rings under study more evident. For the reader interested in a more in-depth description of these notions, we suggest consulting the books by Hall [49] and Grilliet [48], together with a more specialized reference such as the book on finite representation theory by Burrow [21].

We note that all finite groups are linear, as a consequence of the Cayley representation theorem and the following property:

Proposition 1.47 *Let $\mathbb{K}$ be any ring and S be the set of all $d \times d$ matrices from $\mathbb{K}$ with entries in $\{0, 1\}$ which have exactly one 1 on every row and column. Then S is a subgroup of $\mathrm{GL}_d(\mathbb{K})$, isomorphic to the permutation group S_d.*

A matrix with the aforementioned property is called a **permutation matrix**. Of course, this is an "inefficient" matrix representation in the sense of requiring extremely large matrices for large values of $|G|$. By choosing the ring $\mathbb{K}$ and the representative matrices appropriately, we may find representations which require matrices that are significantly smaller than $|G|$; for example, the quaternion group defined below would require 8×8 matrices for a representation of this type, but there exists a natural representation that uses 2×2 matrices over $\mathbb{C}$.

Permutation matrices give an upper bound on the least d for which there exists a d-dimensional linear representation of a finite group G over $\mathbb{Z}$, which is $d \leq |G|$. The following result is useful in estabilishing a lower bound for this d:

Theorem 1.48 *Every finite subgroup of* $\mathrm{GL}_d(\mathbb{Z})$ *is isomorphic to a subgroup of* $\mathrm{GL}_d(\mathbb{Z}/3\mathbb{Z})$*.*

Corollary 1.49 *For every* $d \geq 1$*, there are finitely many isomorphism classes of finite subgroups of* $\mathrm{GL}_d(\mathbb{Z})$*.*

The field $\mathbb{Z}/3\mathbb{Z}$ is finite, and thus, $\mathrm{GL}_d(\mathbb{Z}/3\mathbb{Z})$ must be finite as well. Using basic techniques from combinatorics to determine the ways in which a $d \times d$ matrix can have d linearly independent columns, we get the following result:

Proposition 1.50 $|\mathrm{GL}_d(\mathbb{Z}/3\mathbb{Z})| = \prod_{k=0}^{d-1}(3^d - 3^k)$*.*

Corollary 1.51 *If* $|G| > \prod_{k=0}^{d-1}(3^d - 3^k)$*, then* G *cannot have a d-dimensional linear representation.*

1.6.3. The quaternion group

The construction used to obtain the field of complex numbers $\mathbb{C}$ from the field of real numbers $\mathbb{R}$ by adding an imaginary unit i is quite well known. Slightly lesser known but about just as important is the fact that one may obtain a larger division ring $\mathbb{H}$ (known as the **quaternion skew field**) by adjoining three units instead of one, only losing the commutativity of the multiplication, using a variant of the same construction (which is a subcase of the group ring method from representation theory).

In what follows we are interested in the four units $\{1, \mathrm{i}, \mathrm{j}, \mathrm{k}\}$ which generate $\mathbb{H}$, which generate an eight-element group [49] with the following multiplication table:

	1	i	j	k	−1	−i	−j	−k
1	1	i	j	k	−1	−i	−j	−k
i	i	−1	k	−j	−i	1	−k	j
j	j	−k	−1	i	−j	k	1	−i
k	k	j	−i	−1	−k	−j	i	1
−1	−1	−i	−j	−k	1	i	j	k
−i	−i	1	−k	j	i	−1	k	−j
−j	−j	k	1	−i	j	−k	−1	i
−k	−k	−j	i	1	k	j	−i	−1

We write $Q = \{1, \mathrm{i}, \mathrm{j}, \mathrm{k}, -1, -\mathrm{i}, -\mathrm{j}, -\mathrm{k}\}$ and call it the **quaternion group**. It is not hard to see from the table that Q is a rank 2 group generated by $\{\mathrm{i}, \mathrm{j}\}$, and its elements satisfy the following identities:

$$\mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = \mathrm{ijk} = -1, \tag{1.2}$$

and that $Z(Q) = \{1, -1\} \cong \mathbb{Z}/2\mathbb{Z}$. This is also an example of a group extension that is not a semidirect product, as $\langle \mathrm{i} \rangle \cong \mathbb{Z}/4\mathbb{Z}$ (the group of units that generate $\mathbb{C}$ from $\mathbb{R}$ as a group ring) is a normal order 4 subgroup of Q, and thus $Q/\langle i \rangle \cong \mathbb{Z}/2\mathbb{Z}$; however, the only order 2 subgroup of Q is $Z(Q)$, which is already contained in $\langle \mathrm{i} \rangle$; this shows that there is no group action $\mathbb{Z}/2\mathbb{Z} \overset{\varphi}{\curvearrowright} \mathbb{Z}/4\mathbb{Z}$ such that $Q \cong (\mathbb{Z}/4\mathbb{Z}) \rtimes_{\varphi} (\mathbb{Z}/2\mathbb{Z})$ (see the note on internal semidirect products in the previous section). Furthermore, it can be verified that every subgroup of Q is normal[13] and any nontrivial subgroup of Q contains $Z(G)$

We are interested in the quaternion group as a source of examples for some constructions below. Thus, we list a few more properties in this section.

Proposition 1.52 $\mathrm{Inn}(Q) \cong Q/Z(Q) \cong (\mathbb{Z}/2\mathbb{Z})^2$.

Proposition 1.53 $\mathrm{Aut}(Q) \cong S_4$, *and thus* $\mathrm{Out}(Q) \cong S_3$.

Proposition 1.54 *Q can be represented as a matrix group comprised of 2×2 complex matrices, as it is isomorphic to the subgroup of $\mathrm{GL}_2(\mathbb{C})$ given by:*

$$M_{\mathrm{i}} = \begin{bmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{bmatrix}, \quad M_{\mathrm{j}} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

where i *in the matrix entries is to be taken as $\sqrt{-1} \in \mathbb{C}$ and not the corresponding element of $\mathbb{H}$, to prevent confusion.*

Similarly, Q is isomorphic to the subgroup of $\mathrm{GL}_4(\mathbb{R})$ given by:

$$M'_{\mathrm{i}} = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}, \quad M'_{\mathrm{j}} = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Proposition 1.55 *Q has the following* ***group presentation****:*

$$Q = \langle \mathrm{i}, \mathrm{j} \ : \ \mathrm{i}^4 = 1, \mathrm{i}^2 = \mathrm{j}^2, \mathrm{j}^{-1}\mathrm{i}\mathrm{j} = \mathrm{i}^{-1} \rangle.$$

This means that Q is the "largest" group generated by two elements i *and* j *which satisfies these three identities, as there is a group epimorphism from Q to any other group of rank* 2 *that satisfies these equalities.*

1.6.4. Dihedral groups and other geometrical groups

Remember that an **isometry**[14] of the Euclidean space $\mathbb{R}^d$ is a function $f\colon \mathbb{R}^d \to \mathbb{R}^d$ that preserves distances, that is:

$$(\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d)\colon \|f(\boldsymbol{x}) - f(\boldsymbol{y})\| = \|\boldsymbol{x} - \boldsymbol{y}\|.$$

It is not hard to see that if f fixes a point O in Euclidean d-space, then f is entirely determined by its image on an orthonormal basis of $\mathbb{R}^d$ after choosing a coordinate system with O as

[13] A group with this property is called a **Hamiltonian group**. Every Hamiltonian group contains a copy of Q as a subgroup.

[14] With the definition given here, isometries in $\mathbb{R}^d$ are automatically bijective. Some authors allow the word isometry to refer to non-surjective mappings which preserve distances, such as the natural embedding of $\mathbb{R}^d$ into $\mathbb{R}^{d+1}$ or the shift map from a separable Hilbert space into itself; however, we will not delve into such situations, so our isometries will be bijective.

its point of origin. The image of such a basis under f is again an orthonormal basis (as a consequence of the polarization identities[15]). This implies that f, with regards to this coordinate system, is a linear map and its representative matrix has orthogonal columns (i.e. it is an element of $\mathrm{O}_d(\mathbb{R})$). From this, we can deduce that any isometry of $\mathbb{R}^d$ is of the form:

$$f(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b}, \quad A \in \mathrm{O}_d(\mathbb{R}), \boldsymbol{b} \in \mathbb{R}^d,$$

and, reciprocally, any function of this form is an isometry; thus, isometries form a group which is isomorphic to $\mathbb{R}^d \rtimes_\varphi \mathrm{O}_d(\mathbb{R})$, where φ is the restriction to $\mathrm{O}_d(\mathbb{R})$ of the natural action of $\mathrm{GL}_d(\mathbb{R})$ on $\mathbb{R}^d$.

Let $X \subseteq \mathbb{R}^d$ be any set of points in the plane. We are interested in the set of isometries that preserve X:

Definition 1.56 *The* ***group of rigid symmetries*** *of* X, $\mathrm{Sym}(X)$, *is the set of all isometries* $f\colon \mathbb{R}^d \to \mathbb{R}^d$ *which map* X *to itself, i.e.:*

$$\mathrm{Sym}(X) := \{f\colon \mathbb{R}^d \to \mathbb{R}^d \;:\; f(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b}, A \in \mathrm{O}_d(\mathbb{R}), \boldsymbol{b} \in \mathbb{R}^d, f[X] = X\}.$$

Remark Note that $\mathrm{Sym}(X)$ depends implicitly on the ambient space $\mathbb{R}^n$. For instance, if $X = [0,1] \cup \{-1\}$, there are no nontrivial isometries of $\mathbb{R}$ that map X to itself and thus $\mathrm{Sym}(X) = \{1\}$. However, if we identify $\mathbb{R}$ with the horizontal line $\mathbb{R} \times \{0\} \subset \mathbb{R}^2$ we see that the reflection $m\colon \mathbb{R}^2 \to \mathbb{R}^2$ along the horizontal axis maps X to itself, and thus $\mathrm{Sym}(X)$ has a nontrivial element. To avoid this issue, we usually require $\mathrm{Sym}(X)$ to act faithfully on X, and thus we identify any two symmetries f, g which satisfy $f|_X = g|_X$.

If X is bounded, then there must exist some point O in Euclidean space (not necessarily belonging to X) which is left fixed by all $f \in \mathrm{Sym}(X)$. For instance, if X is measurable and has positive measure, then the centroid of X may be taken as such a point. Thus:

Proposition 1.57 *If* $X \subset \mathbb{R}^d$ *is bounded,* $\mathrm{Sym}(X)$ *is (isomorphic to) a subgroup of* $\mathrm{O}_d(\mathbb{R})$.

Thus, $\mathrm{Sym}(X)$ is a linear group with a d-dimensional representation[16]. We are interested in certain groups of rigid symmetries we shall encounter often in what follows:

Definition 1.58 *For any* $n \geq 3$, *the* ***dihedral group*** D_n *is the group of rigid symmetries of an* n*-sided regular polygon in* $\mathbb{R}^2$. *Alternatively*[17]*:*

$$D_n := \mathrm{Sym}(\{e^{2\pi i k/n} \;:\; 0 \leq k < n\}).$$

It is not hard to see that D_n is generated by a rotation r by an angle of $2\pi/n$ and a reflection m through any axis of symmetry; using this, it is easy to see that:

[15] Which allow us to write the inner product in $\mathbb{R}^d$ in terms of the norm.

[16] Using some techniques from linear algebra, we can see that in the general case $\mathrm{Sym}(X)$ is also a linear group, having a $(d+1)$-dimensional representation.

[17] This allows us to define D_2 as well, which corresponds to the Klein 4-group $(\mathbb{Z}/2\mathbb{Z})^2$ and may be thought of as the set of rigid symmetries of a rectangle. Note that in this case D_2 does not act faithfully on $\{+1, -1\} \subset \mathbb{C}$, but by using this convention for D_2 the semidirect product and group presentation schema of the other dihedral groups is still satisfied.

Proposition 1.59 *$D_n \cong (\mathbb{Z}/n\mathbb{Z}) \rtimes (\mathbb{Z}/2\mathbb{Z})$ (where the group action is the only possible nontrivial one), and thus it has the following group presentation:*

$$D_n = \langle r, m \,:\, r^n = 1, m^2 = 1, (rm)^2 = 1\rangle.$$

As well, we have the following matrix representation:

Proposition 1.60 *D_n is isomorphic to the subgroup of $\mathrm{GL}_2(\mathbb{R})$ given by the two matrices:*

$$R = \begin{bmatrix} \cos(2\pi/n) & -\sin(2\pi/n) \\ \sin(2\pi/n) & -\cos(2\pi/n) \end{bmatrix}, \quad M = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Another group that appears fairly often is the following, which is a sort of "limit of D_n when n tends to infinity":

Definition 1.61 *The **infinite dihedral group** D_∞ corresponds to $\mathrm{Sym}(\mathbb{Z})$ (with $\mathbb{R}$ as ambient space).*

Proposition 1.62 *As in the case of finite dihedral groups, D_∞ is generated by the translation $r(n) = n + 1$ and the reflection $m(n) = -n$ or, alternatively, by the two reflections m and $s(n) = 1 - n$. Thus, it has the following two group presentations:*

$$\begin{aligned} D_\infty &= \langle r, m \,:\, m^2 = 1, (rm)^2 = 1\rangle, \\ &= \langle s, m \,:\, m^2 = 1, s^2 = 1\rangle. \end{aligned}$$

It has a 2-dimensional real (integral) linear representation generated by the two matrices:

$$R = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad M = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

D_∞ is isomorphic to a semidirect product of the form $\mathbb{Z} \rtimes (\mathbb{Z}/2\mathbb{Z})$, where the group action is the only possible nontrivial one.

Finally, while we have already addressed these groups as an example of wreath products, it is worth mentioning a few properties of the groups of symmetries of d-dimensional cubes. These groups are described thoroughly by Baake [2].

Definition 1.63 *The d-dimensional **hyperoctahedral group** Q_d is the group of rigid symmetries of a d-dimensional cube, that is, $\mathrm{Sym}([-1,1]^d)$.*

For instance, $Q_2 = D_4$, the group of symmetries of a square. As stated before, this group has the following structure:

Proposition 1.64 *We have $Q_d \cong (\mathbb{Z}/2\mathbb{Z}) \wr S_d = (\mathbb{Z}/2\mathbb{Z})^d \rtimes S_d$, and thus $|Q_d| = 2^d d!$. This group has a d-dimensional real matrix representation given by the set of all **signed permutation matrices**, that is, matrices with a single nonzero entry on every row and column, which may be either 1 or -1.*

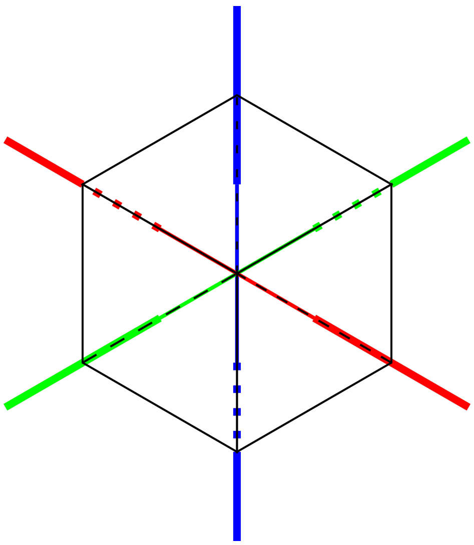


Figure 1.1: The three-dimensional hyperoctahedral group Q_3 as the group of rigid symmetries of a cube. Seen as a wreath product, the S_3 component corresponds to permutations of the coordinate axes shown in the figure, irregardless of orientation; the $\mathbb{Z}/2\mathbb{Z}$ components represent reflections along the plane orthogonal to the corresponding axis.

Every signed permutation matrix is a product of a permutation matrix and a diagonal matrix whose diagonal entries are either 1 or -1. This matrix representation can be derived from the fact that any $A \in Q_d$ must permute the coordinate axes (corresponding to the permutation matrix) and then might change the orientation of some of them (depending on the signs of the corresponding diagonal matrix).

1.6.5. Groups of p-adic integers and odometers

We conclude this section with a brief description of a family of infinite groups we shall encounter fairly often in the following chapters. In what follows, $p > 1$ will be a fixed integer[18]. First, note that if $m \equiv n \pmod{p^k}$, then $m \equiv n \pmod{p^\ell}$ for every $\ell < k$. This defines naturally a surjective group morphism $p_{\ell,k} \colon \mathbb{Z}/p^k\mathbb{Z} \to \mathbb{Z}/p^\ell\mathbb{Z}$ for every pair of positive integers $\ell < k$ (which is actually the quotient morphism for the group $(\mathbb{Z}/p^k\mathbb{Z})/(p^\ell\mathbb{Z}/p^k\mathbb{Z})$).

The collection of all cyclic groups $Z/p^k\mathbb{Z}$, together with the mappings $p_{\ell,k}$, form an inverse system, as defined above. Thus, their projective limit is well-defined, and equals the following group:

Definition 1.65 *Given any integer $p > 0$, the* ***group of p-adic integers***[19] *is the inverse*

[18] In most applications, p is taken to be a prime. However, for our purposes, it is useful to allow p to be a composite integer.

[19] Actually, multiplication is also well-defined for p-adic integers, defined componentwise just like addition, and thus $\mathbb{Z}_p$ is actually a ring. However, if p is composite, $\mathbb{Z}_p$ contains zero divisors, which is why in applications where the multiplicative structure matters p is usually expected to be a prime. In this work, we are only interested in the group structure.

limit given by:

$$\mathbb{Z}_p := \lim_{\longleftarrow k} \mathbb{Z}/p^k\mathbb{Z} = \left\{ (n_k)_{k\geq 1} \in \prod_{k=1}^{\infty} \mathbb{Z}/p^k\mathbb{Z} \ : \ (\forall k > \ell) \colon n_k \equiv n_\ell \pmod{p^\ell} \right\}.$$

That is, every element of $\mathbb{Z}_p$ may be thought of as a sequence of positive numbers $n_1, n_2, \ldots$, such that every n_k leaves a remainder of n_{k-1} when divided by p^{k-1}. If we write the numbers $n_1, n_2, \ldots$ in base p, we see that this means that each n_ℓ corresponds to the last ℓ digits of every n_k that appears afterwards. For example, the following would be a typical element of $\mathbb{Z}_{10}$:

$$x = (5, 5, 5, 3005, 43005, 43005, 8043005, \ldots)$$

Addition is performed componentwise. Alternatively, one may think of an element of $\mathbb{Z}_p$ as a sequence of digits $d_0, d_1, d_2, \ldots$ between 0 and $p-1$, extending infinitely to the left, such that $p^k d_k + n_k = n_{k+1}$, i.e. $n_1 = d_0, n_2 = pd_1 + d_0$, and in general:

$$n_k = d_0 + pd_1 + p^2 d_2 + \ldots + p^{k-1} d_{k-1}.$$

For instance, the aforementioned x would correspond to a sequence of the form $\ldots 8043005$ in this representation. In this case, addition is performed using the "carry the one" algorithm used to compute the sum of two numbers in base p in $\mathbb{Z}$, repeated indefinitely to the left.

The group $\mathbb{Z}_p$ naturally contains an embedded copy of $\mathbb{Z}$: for any positive n, there exists an element $(n_1, n_2, \ldots) \in \mathbb{Z}$ that eventually stabilizes at n (i.e. $n_k = n$ for all $k \geq k_0$) and it is not hard to see that addition of any two such sequences stabilizing at n and m results in a sequence stabilizing at $n + m$ (in particular, $(1, 1, 1, 1, \ldots)$ is the generating element for a copy of $\mathbb{N}$ inside $\mathbb{Z}_p$ which contains all eventually stabilized sequences). In terms of the digits $d_0, d_1, \ldots$, this corresponds to taking the base p representation of n and extending it with infinitely many zeros to the left. Similarly, we may think of elements which have infinitely many digits $p-1$ to the left as negative numbers; for instance, the decimal expansion of -1 in $\mathbb{Z}_{10}$ is $\ldots 9999999$. We identify $\mathbb{Z}$ with this subgroup of $\mathbb{Z}_p$ and make no difference between, e.g. the integer 104 and the 10-adic integer $\ldots 000000104$, nor between $\ldots 99999911$ and -89. Note, though, that $\mathbb{Z}_p$ is always strictly larger than $\mathbb{Z}$; for instance, the infinite sequence of digits $\ldots 101010101$ does not correspond to any integer.

The group $\mathbb{Z}_p$ can be given a **topological group** structure, i.e. it can be given a topology that makes the operations of addition $(x, y) \mapsto x + y$ and inverses $x \mapsto -x$ continuous. It is actually a metric space, via the following distance function:

$$K(\ldots d_3 d_2 d_1 d_0, \ldots e_3 e_2 e_1 e_0) = \begin{cases} N & \text{if } d_N \neq e_N \wedge d_k = e_k \text{ for all } k < N, \\ \infty & \text{if } d = k = e_k \text{ for all } k, \end{cases}$$

$$\delta_p(x, y) = p^{-K(x,y)}.$$

This is a particular case of the **shift metric** we shall encounter in a latter chapter; we see that $\delta_p(x, y) < p^{-n}$ if x and y match on their first n digits, read from right to left. Note that under this metric sequences such as p^k and $1 + p + p^2 + \ldots + p^k$ converge; e.g., in $\mathbb{Z}_{10}$ the sequence $9, 99, 999, \ldots$ converges to -1. Furthermore, $\mathbb{Z}$ is a dense subset of $\mathbb{Z}_p$ in this topology, as if $x = (n_1, n_2, \ldots)$, then $\delta_p(n_k, x) \xrightarrow{k\to\infty} 0$.

The following two results allow us to understand the structure of $\mathbb{Z}_p$ when p is not a prime:

Proposition 1.66 *For any $k \geq 1$ and prime p, we have $\mathbb{Z}_p \cong \mathbb{Z}_{p^k}$.*

Proposition 1.67 *If $\gcd(p,q) = 1$, then $\mathbb{Z}_{pq} \cong \mathbb{Z}_p \times \mathbb{Z}_q$.*

Since $\mathbb{Z}_p$ is a topological group which contains a copy of $\mathbb{Z}$, one may define a group action $\mathbb{Z} \overset{\omega}{\curvearrowright} \mathbb{Z}_p$, totally determined by the function $\omega\colon \mathbb{Z}_p \to \mathbb{Z}_p$ given by $\omega(x) = x+1$. The function ω is continuous, and thus, the pair $(\mathbb{Z}_p, \omega)$ is a topological dynamical system, called a p-adic **odometer**. Since the orbit of any x is $x + \mathbb{Z}$, all p-adic odometers are minimal; furthermore, they have other interesting dynamical properties such as equicontinuity and zero entropy, and they appear often in relation to more complicated systems.

Similarly to the one-dimensional case, we may define d-dimensional analogues of the p-adic integers (and thus of p-adic odometers) by taking any matrix U with integer entries and satisfying the condition $|\det(U)| > 1$, and defining an inverse system using the natural maps $p_{\ell,k}\colon \mathbb{Z}^d/U^k\mathbb{Z}^d \to \mathbb{Z}^d/U^\ell\mathbb{Z}^d$. When U is a diagonal matrix whose entries are positive integers $s_1, \dots, s_d > 1$, the resulting projective limit group $\mathbb{Z}_U$ is isomorphic to the product $\mathbb{Z}_{s_1} \times \cdots \times \mathbb{Z}_{s_d}$, but in the general case no such decomposition exists. Regardless, we obtain a topological group once again, with a similar shift metric and a natural embedding of $\mathbb{Z}^d$ as a dense subset, and thus there is a minimal group action $\mathbb{Z}^d \overset{\omega}{\curvearrowright} \mathbb{Z}_U$ with nice properties. We shall encounter some of these d-dimensional odometers in latter chapters.

Chapter 2

Algebraic number theory and algebraic geometry

One of our subjects of interest in what follows is the setting of subshifts defined via algebraic conditions. In this line of work, we need some basic concepts from ring theory, specially from a number-theoretic viewpoint, which we proceed to detail below. Afterwards, we detail some very basic notions of algebraic geometry and the theory of algebraic sets.

2.1. Basic notions of rings

Definitions below can be consulted in most abstract algebra references, such as the aforementioned book by Grilliet [48], or in specialized books on algebraic number theory, such as Jarvis [55] or Neukirch [81]. Some basic knowledge on general number theory is also advised, although not required; the author recommends the introductory book from Projeto Euclides [76].

Definition 2.1 *A* ***ring*** *is an ordered triple $(R, +, \cdot)$ where R is a set and both $+$ (the* ***ring addition****) and $\cdot$ (****ring multiplication****) are functions $R \times R \to R$ satisfying the following properties:*

- *$(R, +)$ is an abelian group with neutral element 0_R. Inverses in this context are denoted via additive notation, i.e. the inverse of $x \in R$ is written as $-x$.*
- *$(R, \cdot)$ is a* ***monoid****, i.e. $\cdot$ is an associative (but not necessarily commutative) operation with neutral (identity) element 1_R. Note that we do not expect all elements of R to have a multiplicative inverse. When they do exist, we use multiplicative notation and write x^{-1} for the inverse of x, and in this case we say that x is a* ***unit****. The set $R^\times$ of all units is a group under multiplication.*
- *The operation $\cdot$ is* ***distributive*** *over $+$, i.e. it satisfies the following two equalities:*

$$\begin{aligned}(\forall x, y, z \in R)\colon x \cdot (y + z) &= (x \cdot y) + (x \cdot z),\\ (x + y) \cdot z &= (x \cdot z) + (y \cdot z).\end{aligned}$$

*When the operation $\cdot$ is commutative, we say that R is a **commutative ring** or **domain**[1]. If all nonzero elements of $R \setminus \{0_R\}$ have a multiplicative inverse, we say that R is a **division ring**. A commutative division ring is called a **field**.*

Examples of commutative rings include the integers $\mathbb{Z}$, the rationals $\mathbb{Q}$, the reals $\mathbb{R}$ and the complex numbers $\mathbb{C}$, each with their standard addition and multiplication operations. The latter three are also fields, although the first is not. The set of all $d \times d$ matrices, $\mathbb{M}_d(\mathbb{R})$, is a noncommutative ring, which exhibits the following phenomenon:

Definition 2.2 *Let R be a ring. We say R has **zero divisors** if there exist $x, y \in R \setminus \{0_R\}$ such that $x \cdot y = 0_R$. A commutative ring without zero divisors is called an **integral domain**.*

We shall be interested mostly in integral domains, although some rings with zero divisors can make appearances sometimes, such as the aforementioned matrix ring $\mathbb{M}_d(\mathbb{R})$. We note that the general linear group introduced previously, $\mathrm{GL}_d(\mathbb{R})$, corresponds exactly to the group of units of $\mathbb{M}_d(\mathbb{R})$.

Several of the usual definitions from group theory carry over to rings with minimal changes. However, some issues arise that are important to keep in mind, so we make sure to note what things change.

Definition 2.3 *Let R be a ring. A subset $S \subseteq R$ is a **subring** of R if:*

- *it is a ring as well under the same operations $+$ and $\cdot$, restricted to S (i.e. $(S, +)$ is an abelian subgroup of $(R, +)$, the product of two elements of S is in S, etc.), and*
- *$1_R \in S$, that is, they have the same multiplicative identity.*

*An **ideal**[2] of R is an additive subgroup $(\mathfrak{i}, +)$ of $(R, +)$ that satisfies the following property:*

$$r \in R, a \in \mathfrak{i} \implies ra, ar \in \mathfrak{i}.$$

Remark The requirement of 1_R belonging to any subring is non-obvious, and indeed, if R has zero divisors, there may be subsets $S \subset R$ closed under addition and multiplication and that may be rings in their own, but that have a different identity element. For example, we may embed the ring $\mathbb{M}_d(\mathbb{R})$ into $\mathbb{M}_{d+1}(\mathbb{R})$ by filling the additional row and column with zeros, and thus the matrix $\mathrm{diag}(1, \ldots, 1, 0)$ acts as an identity on $\mathbb{M}_d(\mathbb{R})$ under matrix multiplication; however, this is not a true subring of $\mathbb{M}_{d+1}(\mathbb{R})$ as this is not the usual identity matrix.

Note that, while both ideals and subrings are closed under addition and multiplication, an ideal need not to be a subring[3] and vice versa. In fact, if $\mathfrak{i}$ is both a subring and an ideal, the fact that $1_R \in \mathfrak{i}$ implies that $R \subseteq \mathfrak{i}$, since every $r \in R$ equals $r1_R$ and thus also belongs to

[1] Some books have a looser definition of ring, which allows a ring to lack a multiplicative neutral element. Under such a definition, triples such as $(2\mathbb{Z}, +, \cdot)$ would be rings. The word "domain" refers specifically to a commutative ring that has a multiplicative neutral element.

[2] More precisely, a two-sided ideal. We shall mostly deal with commutative rings, so we will not discuss left or right ideals.

[3] Some authors refer to any subset closed under addition and multiplication as a subring, and, with such a definition, ideals would be subrings. However, our additional requirement that 1_R belongs to every subring excludes this situation.

$\mathfrak{i}$. Thus $\mathfrak{i} = R$. Hence, we say that an ideal is nontrivial if it does not equal $\{0\}$ nor R, and that nontrivial ideals are never subrings. We note, as well, that fields do not have nontrivial ideals, as if $x \in \mathfrak{i} \setminus \{0\}$, then $x^{-1}x = 1_R \in \mathfrak{i}$. In general, we have the following:

Proposition 2.4 *If $\mathfrak{i}$ is a nontrivial ideal in R, then $\mathfrak{i} \cap R^\times = \varnothing$.*

While ideals and subrings are now different kinds of objects, they will take the role normal and non-normal subgroups had in group theory. In fact, we have the following:

Proposition 2.5 *Given a nontrivial ideal $\mathfrak{i} \subset R$, the relation $x \sim y$ if $x - y \in \mathfrak{i}$ is an equivalence relation that satisfies the following properties:*

$$x \sim x' \wedge y \sim y' \implies x + y \sim x' + y' \wedge xy \sim x'y'.$$

Thus, the set of all equivalence classes $R/\mathfrak{i}$ is a ring under the operations $[x] + [y] = [x + y]$ and $[x] \cdot [y] = [xy]$, called the ***quotient ring*** *of R over $\mathfrak{i}$.*

Ring homomorphisms, just like their group analogues, are defined as to preserve the structure of the corresponding rings as well as possible, which includes both operations *and* the multiplicative identities:

Definition 2.6 *Let $(R, +, \cdot)$ and $(S, +, \cdot)$ be two rings. A function $f \colon R \to S$ is a* ***ring homomorphism*** *if it satisfies the following properties:*

- $f(x + y) = f(x) + f(y)$,
- $f(xy) = f(x)f(y)$,
- $f(1_R) = 1_S$.

The terms monomorphism, epimorphism and isomorphism retain their meaning from group theory. As well, we think of two isomorphic rings as the same ring for all intents and purposes. The following definition appears very frequently as well:

Definition 2.7 *Given a ring homomorphism $f \colon R \to S$, the* ***kernel*** *of f is the ideal $\mathfrak{k} = \ker(f)$ containing all $r \in R$ with $f(r) = 0_S$.*

It is not hard to check that $f(0_R) = 0_S$ and $0_S \cdot x = 0_S$ for any $x \in S$, in any ring, thus $\mathfrak{k}$ is clearly an ideal, which is the trivial ideal $\{0\}$ if, and only if, f is injective. Once again, the first isomorphism theorem holds:

Theorem 2.8 (First Isomorphism Theorem for rings) *For any ring R and any ring homomorphism $f \colon R \to S$, we have $R/\ker(f) \cong \operatorname{im}(f)$.*

Furthermore, it is worth noting that, since homomorphisms preserve the multiplicative units, we must have $f(x^{-1}) = f(x)^{-1}$ for any $x \in R^\times$, i.e. if $f \colon R \to S$ is a ring homomorphism, f maps $R^\times$ to $S^\times$. In particular, if R is a field, the only element that can be mapped to 0_S is 0_R, and thus f is injective.

In applications to number theory, there are several kinds of ideals that take a central role, as well as many properties of them. We start with a few operations on ideals:

Proposition 2.9 *Given any collection of ideals $\{\mathfrak{a}_i\}_{i\in I}$, their intersection $\bigcap_{i\in I}\mathfrak{a}_i$ is once again an ideal. Thus, given any subset $S\subset R$, we may speak of the* ***ideal generated by*** *S as the following ideal of R:*

$$(S) := \bigcap_{S\subseteq\mathfrak{i}} \mathfrak{i}.$$

The ***product of two ideals*** *$\mathfrak{a}$ and $\mathfrak{b}$ is defined by:*

$$\mathfrak{a}\cdot\mathfrak{b} := (\{ab \;:\; a\in\mathfrak{a}, b\in\mathfrak{b}\}),$$

while their ***sum*** *is given by:*

$$\mathfrak{a}+\mathfrak{b} := \{a+b \;:\; a\in\mathfrak{a}, b\in\mathfrak{b}\} = (\mathfrak{a}\cup\mathfrak{b}).$$

The operations of sum and product of ideals are associative, so we may extend them to any number of ideals. Furthermore, the following inclusions hold:

$$\mathfrak{a}\cdot\mathfrak{b}\subseteq\mathfrak{a}\cap\mathfrak{b}\subseteq\mathfrak{a},\mathfrak{b}\subseteq\mathfrak{a}+\mathfrak{b}.$$

For instance, in $\mathbb{Z}$ all ideals are of the form $(n)=n\mathbb{Z}$. We have that $(m)+(n)=(\gcd(m,n))$, while $(m)\cdot(n)=(mn)$; in contrast, $(m)\cap(n)=(\operatorname{lcm}(m,n))$, with the equality $(m)\cap(n)=(mn)$ holding only when m and n are **coprime**, that is, they have no common factors. In fact, the intuitive interpretation we shall give to ideals is exactly as sets of multiples of a given number (and thus we interpret the sentence $a\in(b)$ as a synonym for $b\mid a$, i.e. a is a multiple of b), and if the ring is “well behaved” this is exactly what they are:

Definition 2.10 *An integral domain $(R,+,\cdot)$ is called a* ***principal ideal domain*** *(PID) if every ideal is generated by a single element, i.e. every ideal is of the form (r) for some $r\in R$. An ideal generated by a single element is called a* ***principal ideal****.*

The ring of integers $\mathbb{Z}$ is the quintaessential example of a principal ideal domain. Fields are PIDs by default, as they do not have any ideals besides the trivial ones (generated by 0_R and 1_R, and thus principal). A less trivial class of PIDs is given by the following fact, which we shall implicitly use often:

Theorem 2.11 *Let K be a field. Then, the ring of polynomials on one variable over K, $K[x]$, is a principal ideal domain.*

However, we often have to deal with rings that do not have such good behavior. For instance, the above theorem does not extend to polynomial rings over several variables, or over rings that are not fields, as both $\mathbb{R}[x,y]$ and $\mathbb{Z}[x]$ are not PIDs. There is a weaker property that has more “stability” in this sense, and we shall make use of it in certain situations:

Definition 2.12 *An integral domain $(R,+,\cdot)$ is called a* ***Noetherian ring*** *if every ideal is* ***finitely generated****, i.e. for every ideal $\mathfrak{i}$ there exist $r_1,\ldots,r_k\in R$ such that $\mathfrak{i}=(r_1,\ldots,r_k)$.*

Theorem 2.13 *A ring R is Noetherian if, and only if*[4] *it satisfies the* ***ascending chain condition****: if $\mathfrak{i}_0\subseteq\mathfrak{i}_1\subseteq\mathfrak{i}_2\subseteq\cdots$ is an ascending sequence of ideals in R, there exists some $n_0\in\mathbb{N}$ where the sequence stabilizes, i.e., for every $n\geq n_0$, we have $\mathfrak{i}_n=\mathfrak{i}_{n_0}$.*

[4] The proof of this equivalence uses the Axiom of Choice.

The following two properties imply that most (if not all) of the rings we will encounter in what follows are Noetherian:

Theorem 2.14 *If R is a Noetherian ring, then $R[x]$, the set of all polynomials with coefficients in R, is Noetherian. In particular, given the isomorphism between $R[x_1, \dots, x_n]$ and $(R[x_1, \dots, x_{n-1}])[x_n]$, any ring of polynomials over finitely many variables with coefficients in a Noetherian ring R is Noetherian as well.*

Theorem 2.15 *If R is Noetherian, then, for any ideal $\mathfrak{i}$ in R, the quotient ring $R/\mathfrak{i}$ is Noetherian.*

Over the next sections we shall see that a very powerful tool to study number rings in algebraic number theory is that we can easily construct isomorphisms between them and quotients of polynomial rings. Indeed, we can "add new elements" to a field or ring that satisfy a desired algebraic equation by taking appropriate quotients. The previous results ensure that, if we start with a Noetherian ring, the end result will be Noetherian as well, which, as we shall see, has important number-theoretical implications.

The following definitions will be central for the next sections:

Definition 2.16 *Let R be an integral domain. A **maximal ideal** is a nontrivial ideal $\mathfrak{p}$ such that, if there is some ideal $\mathfrak{i} \supseteq \mathfrak{p}$, then either $\mathfrak{i} = \mathfrak{p}$ or $\mathfrak{i} = R$ (i.e. $\mathfrak{p}$ is maximal under inclusion among nontrivial ideals). Similarly, $\mathfrak{p}$ will be a **prime ideal** if, whenever a product $ab \in \mathfrak{p}$, then either $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$.*

While these two definitions may seem disparate, it turns out that both are generalizations of the idea of "prime number". Indeed, if we interpret $(m) \supseteq (n)$ as $m \mid n$ (as is the case in $\mathbb{Z}$), maximal ideals (p) correspond exactly to those where if $m \mid p$ then either m is $\pm p$ or ± 1, i.e. p is prime. Similarly, it can be proven that p is prime if and only if $p \mid ab$ implies that $p \mid a$ or $p \mid b$, which corresponds to the second condition. Note that primality of an ideal is a property that also translates to products of ideals, i.e. whenever $\mathfrak{a} \cdot \mathfrak{b} \subseteq \mathfrak{p}$, then $\mathfrak{a} \subseteq \mathfrak{p}$ or $\mathfrak{b} \subseteq \mathfrak{p}$. The two definitions are connected by the following result:

Theorem 2.17 *An ideal $\mathfrak{p}$ is maximal if, and only if, $R/\mathfrak{p}$ is a field; it is prime if $R/\mathfrak{p}$ is an integral domain, and vice versa. Thus, a maximal ideal is always prime. When R is a PID, the converse is also true.*

When R is not a PID, the converse of the latter statement might be false.

2.2. Field extensions and algebraic numbers

Remember that a **field extension** $K|F$ is a pair of fields $F \subset K$. Since in this situation K may be seen as a F-vector space, we call the value $[K : F] = \dim_F(K)$ the **degree** of the extension; if $L \supseteq K \supseteq F$ then the degree satisfies the equality $[L : F] = [L : K] \cdot [K : F]$.

Remark In this section and in what follows, we shall assume that all fields involved have characteristic zero and thus have $\mathbb{Q}$ as a subfield, unless stated otherwise.

Definition 2.18 *Given a field extension $K|F$, an element $\alpha \in K$ is an* ***algebraic number*** *over F if there is a polynomial $p \in F[x]$ with coefficients in F such that $p(\alpha) = 0$. Otherwise, we say that x is a* ***transcendental number*** *over F. An* ***algebraic field extension*** *$K|F$ is one where every element of K is algebraic over F.*

Note that, given $\alpha \in K$ algebraic over F, the set of polynomials $p \in F[x]$ such that $p(\alpha) = 0$ is an ideal. Since F is a field, this ideal is principal, i.e. generated by a single polynomial $m_\alpha(x)$, which may be assumed monic. We call this the **minimal polynomial** of α (over F); this is an **irreducible polynomial** over F, in the sense that $m_\alpha(x)$ does not factor into polynomials of strictly smaller degree in $F[x]$. Note that this also implies that the corresponding ideal $(m_\alpha(x))$ is maximal.

Sums, products and quotients of algebraic numbers are algebraic as well. It may be a hard problem to determine the explicit polynomials $m_{\alpha+\beta}(x)$ or $m_{\alpha\beta}(x)$ from the coefficients of the polynomials $m_\alpha(x)$ and $m_\beta(x)$ (although it can be done via matrix manipulations based on the linear representation techniques briefly described further down), so this fact is usually proved via algebraic tricks based on the following two facts:

Proposition 2.19 *If $K|F$ and $L|K$ are algebraic extensions, then $L|F$ is algebraic as well.*

Proposition 2.20 *If $[K : F] < \infty$, then the extension $K|F$ is algebraic.*

Of course, the converse is false; for instance, we could take the set of all complex numbers that are algebraic over $\mathbb{Q}$ and see that they form a field $\mathbb{A}$, but, since $\sqrt[n]{2} \in \mathbb{A}$ for all values of n and these numbers are all linearly independent over $\mathbb{Q}$, we must have $[\mathbb{A} : \mathbb{Q}] = \infty$. However, all algebraic extensions are, in a way, constructed from finite ones, as we shall see below; we need some terminology first.

Let K be any field and $R \subset K$ be a subring (not necessarily a field itself). Suppose S is any subset of K. Write $R[S]$ for the smallest subring (under inclusion) of K that contains both S and R, and $R(S)$ for the smallest subfield of K that contains both S and R, that is:

$$R[S] = \bigcap_{\substack{R,S \subseteq A}}^{A \text{ subring of } K} A, \quad R(S) = \bigcap_{\substack{R,S \subseteq F}}^{F \text{ subfield of } K} F.$$

Just as in the case of ideals, arbitrary intersections of subrings and subfields result in subrings and subfields, respectively; thus, $R[S]$ and $R(S)$ are both well-defined. It is not hard to verify that:

Proposition 2.21 *Let $R \subseteq K$ be any subring of the field K. If $S = \{\alpha\}$, with $\alpha \in K$, then every element of the ring $R[\alpha]$ is of the form $a_0 + a_1\alpha + \ldots + a_m\alpha^m$ with $a_0, \ldots, a_m \in R$. Similarly, any element of $R(\alpha)$ is of the form:*

$$\frac{a_0 + a_1\alpha + \ldots + a_m\alpha^m}{b_0 + b_1\alpha + \ldots + b_n\alpha^n}, a_0, \ldots, a_m, b_0, \ldots, b_n \in R.$$

A similar result holds for any finite $S \subseteq K$, by using the fact that, e.g. $(R[S])[T] = R[S \cup T]$. From this, we get the following consequence:

Proposition 2.22 *Let $K|F$ be a field extension and $\alpha \in K$. The following are equivalent:*

- *α is algebraic over F,*
- *$[F(\alpha) : F] < \infty$,*
- *$F[\alpha] = F(\alpha)$.*

Thus, any element α of an algebraic extension K of F lies in some finite extension of F contained in K. Furthermore, $\deg(m_\alpha(x)) = [F(\alpha) : F]$. An important tool to study field extensions (linked to the last two propositions) is the following:

Lemma 2.23 *Let $E_\alpha \colon F[x] \to K$ be the morphism that maps every polynomial $p(x) = a_0 + a_1x + \ldots + a_nx^n$ in $F[x]$ to its evaluation in α, i.e. the value $a_0 + a_1\alpha + \ldots + a_n\alpha^n$. If α is algebraic over F, then $\ker(E_\alpha) = (m_\alpha(x))$, and thus, we have an isomorphism:*

$$F[x]/(m_\alpha(x)) \cong F(\alpha).$$

Similarly, for any irreducible $p(x) \in F[x]$, the quotient $F[x]/(p(x))$ is a field which is isomorphic to a finite field extension of F of the form $F(\alpha)$, and $p(x) = m_\alpha(x)$.

An example of usage of this lemma is the well-known case of the construction of the complex numbers from the reals, in where we have $\mathbb{C} \cong \mathbb{R}[x]/(x^2+1)$. The equivalence class of x in the latter quotient satisfies the condition $[x]^2 + [1] = [0]$, i.e. $[x]$ takes the role of the imaginary unit i. As we know, the resultant field $\mathbb{C}$ satisfies the property[5] that every polynomial in $\mathbb{C}[x]$ factors into linear terms; this is a particular case of the following:

Definition 2.24 *A field K is* ***algebraically closed*** *if, for every nonconstant polynomial $p \in K[x]$ has a zero in K (and, thus, p factors into a product of linear terms $(ax + b)$). In other terms, K is algebraically closed if every element that would be algebraic over K already belongs to K. The* ***algebraic closure*** *of K is any field $\overline{K}$ satisfying the following two conditions:*

- *$\overline{K}$ is algebraically closed, and*
- *the extension $\overline{K}|K$ is algebraic.*

It can be proved that the algebraic closure of a field always exists and is unique up to isomorphism, so speaking of "the" algebraic closure is completely justified. For example, the aforementioned subfield $\mathbb{A}$ of $\mathbb{C}$ may be proved to be the algebraic closure of $\mathbb{Q}$. Note that $\mathbb{A}$ does not equal $\mathbb{C}$; in particular, the former is denumerable while the latter is not. Since we are interested mostly in algebraic field extensions, we shall assume that the larger field K from a field extension $K|F$ is a subfield of $\overline{F}$ unless stated otherwise.

Before concluding this section, we need to introduce a special sort of ring homomorphism that is closely linked to algebraic field extensions.

Definition 2.25 *Let $K|F$ be a field extension. A* ***F-automorphism*** *of K is a ring automorphism $f \colon K \to K$ which satisfies the property $f(x) = x$ for all $x \in F$. The set of all*

[5] Widely known as the **Fundamental Theorem of Algebra**.

F-automorphisms of K is a group under composition, which we shall denote as $\mathrm{Aut}(F|K)$ *or*[6] $\mathrm{Gal}(F|K)$*.*

For instance, complex conjugation $a+b\mathrm{i} \mapsto a-b\mathrm{i}$ is a $\mathbb{R}$-automorphism of $\mathbb{C}$ (and, as we shall see, the only nontrivial one). Note that, if p is a polynomial with coefficients in F and f is a F-automorphism of K, then $p(f(\alpha)) = f(p(\alpha))$ for every $p \in F[x]$, which means that, in particular, if α is a zero of p, then $f(\alpha)$ is a zero as well. Thus, if $K = F(\alpha)$, f is entirely determined by the permutation it induces on the roots of $m_\alpha(x)$, as long as it maps α to some element of $F(\alpha)$. The latter condition is actually very important, so we give it a name:

Definition 2.26 *An algebraic field extension $K|F$ is a* ***normal extension*** *if, whenever a polynomial $p \in F[x]$ has a root in K, then every other root of p is also in K, that is, p factors as $c(x-\alpha_1)\cdots(x-\alpha_n)$ with $\alpha_1,\ldots,\alpha_n \in K$.*

$K|F$ is called a ***separable extension*** *if every $\alpha \in K$ has a minimal polynomial without multiple roots, i.e. $m_\alpha(x)$ has no factors of the form $(x-\beta)^r$ with $r \geq 2$ after factoring in $\overline{F}$.*

Finally, a ***Galois extension*** *$K|F$ is one that is both normal and separable.*

We actually only care about normal extensions in this context, as separability is automatic for fields with characteristic zero (and thus "normal" and "Galois" are synonyms for our purposes here). An example of a non-normal algebraic extension is $\mathbb{Q}(\sqrt[3]{2})$, as this field is a subfield of $\mathbb{R}$; the minimal polynomial of $\sqrt[3]{2}$, x^3-2, has three distinct roots, but the remaining two are complex and thus cannot belong to $\mathbb{Q}(\sqrt[3]{2})$. Obviously, the algebraic closure of a field is a Galois extension, but it is not the only one:

Definition 2.27 *Let F be a field and $p \in F[x]$ a polynomial. The* ***splitting field*** *of F is the smallest field $\overline{F} \supseteq K \supseteq F$ such that p factors into linear terms in K.*

Proposition 2.28 *If F has characteristic zero, then $K|F$ is a Galois extension of F if, and only if, K is the splitting field of some polynomial $p \in F[x]$.*

Note that the splitting field of a polynomial $p \in F[x]$ is a finite extension of F and, thus, for any field extension $K|F$, there exists some field extension $N|K$ such that $N|F$ is finite and Galois. An important class of examples of Galois extensions we shall encounter later on is the following:

Definition 2.29 *A* ***cyclotomic field*** *is the splitting field of the polynomial $x^n - 1 \in \mathbb{Q}[x]$. If we write $\zeta_n = \mathrm{e}^{2\pi\mathrm{i}/n}$, then the cyclotomic field associated to x^n-1 is precisely $\mathbb{Q}(\zeta_n)$.*

When $K|F$ is a Galois extension, we use the notation $\mathrm{Gal}(F|K)$ to refer to the group of all F-automorphisms of K. The reason is due to the following central result:

Theorem 2.30 (Fundamental Theorem of Galois theory) *If $K|F$ is a finite Galois field extension, there is a 1-1 correspondence between subfields $K \supseteq L \supseteq F$ and subgroups of*

[6] We reserve this second notation for a specific subcase, which will be detailed below.

$\mathrm{Gal}(K|F)$. *More precisely, if $H \leq \mathrm{Gal}(K|F)$, write:*

$$K^H := \{x \in K \ : \ f(x) = x \text{ for every } f \in H\}.$$

Then every subfield $K \supseteq L \supseteq F$ is of the form $L = K^H$ for precisely one subgroup H of $\mathrm{Gal}(K|F)$. *Furthermore, $H > H'$ if, and only if, $K^H \subset K^{H'}$.*

The **Galois group** $\mathrm{Gal}(K|F)$ will be important for some applications in the next sections. We note a few properties of interest:

Proposition 2.31 *If $K|F$ is a finite Galois extension (and thus a splitting field for some $p \in F[x]$) then $|\mathrm{Gal}(K|F)| = [K : F]$. In the non-Galois case, $|\mathrm{Aut}(K|F)|$ divides $[K : F]$.*

Proposition 2.32 $\mathrm{Gal}(\mathbb{Q}(\zeta_n)|\mathbb{Q}) \cong \mathbb{Z}/n\mathbb{Z}$.

Proposition 2.33 *If $[K : \mathbb{Q}] = 2$, then $K|\mathbb{Q}$ is normal and* $\mathrm{Gal}(K|\mathbb{Q}) \cong \mathbb{Z}/2\mathbb{Z}$. *Such a K is called a **quadratic field**.*

Closely related to F-automorphisms (and thus the associated Galois group), but sometimes easier to handle, is the notion of F-embeddings, defined as follows:

Definition 2.34 *Let $K|F$ be a field extension. A **F-embedding** is any ring homomorphism $\iota\colon K \hookrightarrow \overline{F}$ that satisfies $\iota(x) = x$ for every $x \in F$.*

In the previous definition, $\overline{F}$ may be replaced by any algebraically closed field that contains F. Since for our purposes F will usually be $\mathbb{Q}$ (or a finite extension thereof), we may assume that $\mathbb{Q}$-embeddings are ring homomorphisms $\iota\colon K \to \mathbb{C}$, which makes no difference. For instance, $\mathbb{Q}(\sqrt{2})$ has two $\mathbb{Q}$-embeddings into $\mathbb{C}$, the standard inclusion map and the map $\iota(a + b\sqrt{2}) = a - b\sqrt{2} \in \mathbb{C}$, while $\mathbb{Q}(\sqrt[3]{2})$ has three $\mathbb{Q}$-embeddings, which are entirely determined by whether they map $\sqrt[3]{2}$ to any of the three elements $\sqrt[3]{2}, \zeta_3\sqrt[3]{2}, \zeta_3^2\sqrt[3]{2}$. In general, we have the following:

Proposition 2.35 *There are exactly $[K : F] = d$ F-embeddings $\iota\colon K \hookrightarrow \overline{F}$.*

Since we may assume without loss of generality that $\mathbb{Q}$-embeddings of a field extension $K|\mathbb{Q}$ map K to some subfield of $\mathbb{C}$, it makes sense to distinguish whether this subfield is also a subfield of $\mathbb{R}$ or not; therefore, we speak of **real embeddings** $\iota\colon K \hookrightarrow \mathbb{R}$ and conjugate pairs of **complex embeddings** $\iota, \bar{\iota}\colon K \hookrightarrow \mathbb{C}$, depending on the case; if there are r distinct real embeddings and s pairs of conjugate complex embeddings, we must have $r + 2s = d$.

Proposition 2.36 *A field extension $K|F$ is Galois if, and only if, there is some fixed subfield $\tilde{K} \subseteq \overline{K}$ such that the image of K under any F-embedding is $\tilde{K}$. In such a case, given a fixed embedding $\iota\colon K \hookrightarrow \overline{F}$, every other F-embedding is of the form $\iota \circ f$, where $f \in$* $\mathrm{Gal}(K|F)$.

This gives an explanation to the fact that $\mathbb{Q}(\sqrt[3]{2})$ is not Galois, as, while the three fields $\mathbb{Q}(\sqrt[3]{2})$, $\mathbb{Q}(\zeta_3\sqrt[3]{2})$ and $\mathbb{Q}(\zeta_3^2\sqrt[3]{2})$ are isomorphic, they correspond to distinct subsets of $\mathbb{C}$, and the intersection of any two of them is just $\mathbb{Q}$.

Definition 2.37 *Suppose that $K|\mathbb{Q}$ is a finite field extension with r real $\mathbb{Q}$-embeddings $\iota_1, \dots, \iota_r \colon K \hookrightarrow \mathbb{R}$ and s conjugate pairs of complex embeddings $\sigma_1, \bar{\sigma}_1, \dots, \sigma_s, \bar{\sigma}_s \colon K \hookrightarrow \mathbb{C}$. The **Minkowski embedding** of K is the function $M \colon K \to \mathbb{R}^r \times \mathbb{C}^s$ given by:*

$$M(x) = (\iota_1(x), \dots, \iota_r(x), \sigma_1(x), \dots, \sigma_s(x)),$$

where we choose only one of each conjugate pair of complex embeddings.

This embedding (or, more precisely, its restriction to a specific subset that will be defined afterwards) will take an important role in a later chapter.

We say that two elements $\alpha, \beta \in \overline{F}$ (or, in general, in K for a given field extension $K|F$) are **conjugate** if they share the same minimal polynomial $m_\alpha = m_\beta \in F[x]$. We see that two different F-embeddings of K map a given $x \in K$ to conjugate elements of $\overline{F}$; similarly, F-automorphisms from $\mathrm{Aut}(K|F)$ map any element of K to one of its conjugates, and, if $K|F$ is Galois, all conjugates of any $x \in K$ are of the form $f(x)$ for some $f \in \mathrm{Gal}(K|F)$. Since m_α has a finite number of roots, every element of K has finitely many conjugates, and thus we may add or multiply all of them. The resulting elements of K are mapped to the same number $\nu \in \overline{F}$ by any F-embedding of K, and it can be proved that this only happens when $\nu \in F$ (e.g. a number is real if, and only if, it is equal to its complex conjugate). Thus, we give these numbers a name:

Definition 2.38 *Let $K|F$ be a finite field extension and $x \in K$. The **norm** of x, $N_{K|F}(x)$, is the product of all conjugates of x, while the **trace** $\mathrm{tr}_{K|F}(x)$ of x is the sum of these conjugates. In symbols:*

$$N_{K|F}(x) = \prod_{\iota \colon K \hookrightarrow \overline{F}} \iota(x), \quad \mathrm{tr}_{K|F}(x) = \sum_{\iota \colon K \hookrightarrow \overline{F}} \iota(x).$$

While $N_{K|F}(x)$ and $\mathrm{tr}_{K|F}(x)$ are, by definition, elements of $\overline{F}$, it is not hard to verify that they must belong to F itself. Thus, the norm and trace may be thought of as maps $K \to F$ with the following properties:

$$\begin{aligned} N_{K|F}(xy) &= N_{K|F}(x)N_{K|F}(y), & \mathrm{tr}_{K|F}(x+y) &= \mathrm{tr}_{K|F}(x) + \mathrm{tr}_{K|F}(y), \\ N_{K|F}(x) &= x^{[K:F]},\ x \in F, & \mathrm{tr}_{K|F}(x) &= [K:F]x,\ x \in F, \\ N_{L|F}(x) &= N_{L|K} \circ N_{K|F}(x),\ F \subseteq K \subseteq L, & \mathrm{tr}_{L|F}(x) &= \mathrm{tr}_{L|K} \circ \mathrm{tr}_{K|F}(x),\ F \subseteq K \subseteq L. \end{aligned}$$

The term "norm" comes from the fact that, for the field $\mathbb{Q}(\mathrm{i})$, $N_{\mathbb{Q}(\mathrm{i})|\mathbb{Q}}(a+b\mathrm{i}) = a^2+b^2 = |a+b\mathrm{i}|^2$, and early studies on the ring $\mathbb{Z}[\mathrm{i}]$ and the associated field $\mathbb{Q}(\mathrm{i})$ stated many of its results in terms of the standard complex norm. Regardless of this name, the norm may take positive or negative values; for instance, in the field extension $\mathbb{Q}(\sqrt{2})|\mathbb{Q}$, we have:

$$N_{\mathbb{Q}(\sqrt{2}|\mathbb{Q})}(a + b\sqrt{2}) = (a + b\sqrt{2})(a - b\sqrt{2}) = a^2 - 2b^2,$$

and thus, for instance, $N_{\mathbb{Q}(\sqrt{2})|\mathbb{Q}}(\sqrt{2}) = -2$. In the next section we shall see some properties of norms and traces more closely connected to algebraic number theory, which we shall make use of afterwards.

2.3. Algebraic integers and number rings

We are now in a position to introduce the basic notions of algebraic number theory we shall employ in latter chapters. Our main objects of study will be the following:

Definition 2.39 *Let $K|\mathbb{Q}$ be a field extension. An algebraic number $\alpha \in K$ is called an* ***algebraic integer*** *if its minimal polynomial $m_\alpha(x)$ (which, as stated before, is assumed to be monic) has integer coefficients.*

That is, $\alpha \in K$ is an algebraic integer if $p(\alpha) = 0$ for some $p \in \mathbb{Z}[x]$ with leading coefficient 1. Numbers such as $\sqrt{2}$ and $\varphi = (1 + \sqrt{5})/2$ are algebraic integers, but others such as $1/\sqrt{3}$ are not, despite being algebraic numbers.

Definition 2.40 *Given a field extension $K|\mathbb{Q}$, the* ***ring of integers*** *or* ***maximal order***[7] *of K is the set of all algebraic integers of K, written $\mathcal{O}_K$ (or sometimes $\mathbb{Z}_K$).*

The set $\mathcal{O}_K$ is effectively a subring of K, as it should be obvious that $1 \in \mathcal{O}_K$ and this set is closed under addition and multiplication, even if the verification of this is not as obvious as it might appear at first. The name "algebraic integer" comes from the fact that $\mathcal{O}_K$ has a similar relationship with K that $\mathbb{Z}$ has with $\mathbb{Q}$ (and, in fact, $\mathcal{O}_Q = \mathbb{Z}$), up to and including the fact that every element of K is of the form n/m with $n, m \in \mathcal{O}_K$.

Proposition 2.41 *The ring $\mathcal{O}_K$ is an* ***integrally closed*** *subset of K, meaning that if $p(x) \in \mathbb{Z}[x]$ is a monic polynomial with a root on K, that root belongs to $\mathcal{O}_K$.*

Remark In particular, if a monic polynomial with integer coefficients has a rational root, that root must be an integer.

2.3.1. Divisibility and factorization

One of our main subjects of interest in regards to a ring of algebraic integers is a generalized notion of divisibility, which is connected with many questions from number theory. For instance, a classical question is to determine which numbers can be represented as the sum of two squares of integers (e.g. $5 = 1^2 + 2^2$, but it can be verified by hand that 7 does not equal $a^2 + b^2$ for any $a, b \in \mathbb{Z}$). One way to solve this question is to note that the following set of complex numbers:

$$\mathbb{Z}[\mathrm{i}] = \{a + b\mathrm{i} \ : \ a, b \in \mathbb{Z}\},$$

is actually the ring of algebraic integers of the field $\mathbb{Q}(\mathrm{i})$, whose field norm is $N_{\mathbb{Q}(\mathrm{i})|\mathbb{Q}}(a+b\mathrm{i}) = a^2 + b^2$. Thus, to determine whether $n \in \mathbb{N}$ is a sum of two squares, we need to either find a number $u \in \mathbb{Z}[\mathrm{i}]$ with $N_{\mathbb{Q}(\mathrm{i})|\mathbb{Q}}(u) = n$, or show that such a number does not exist. This question actually relates to the factorization of n as an element of $\mathbb{Z}[\mathrm{i}]$: it can be shown that n is a sum of two squares if it is of the form $n = a^2 b$, where $a, b \in \mathbb{N}$ and b is a product of prime numbers who are also sums of two squares. The latter primes are exactly those who

[7] An **order** is a ring all of whose elements are algebraic integers. Thus, the ring of integers of K is an order that is maximal under inclusion. As we shall see, rings such as $\mathbb{Z}[\sqrt{5}]$ are non-maximal orders.

cease being prime in $\mathbb{Z}[\mathrm{i}]$, factoring into smaller terms, showing how this classical problem reduces to an issue of divisibility on a specific ring.

Definition 2.42 *Let R be an integral domain, and $x, y \in R$. We say that x is a **divisor** of y (or x **divides** y) if there exists $z \in R$ such that $xy = z$; in this situation, we write $x \mid y$. If $x \mid y$ and $y \mid x$, we say that x and y are **equivalent**.*

Remark Two elements $x, y \in R$ are equivalent if, and only if, there exists some unit $u \in R^\times$ such that $x = uy$. When we deal with factorizations, we usually disregard the units from $R^\times$, and we treat two factorizations that differ only in the presence or absence of certain units as the same.

Most of our study of divisibility in $\mathbb{Z}$ relies on "splitting" a given number into its component parts, that is, factorization into prime numbers. We would like to reintroduce this tool in the general setting of rings; however, we will encounter several difficulties that do not appear for our usual integers in $\mathbb{Z}$. First and foremost, we note that we may extend the concept of prime number to integral domains in two different, non-equivalent ways:

Definition 2.43 *Let R be an integral domain, and $p \in R$. We say that p is **irreducible** if, whenever $q \mid p$, either $p \mid q$ (that is, q is equivalent to p) or $q \in R^\times$.*

*We say that p is **prime** if, whenever $p \mid ab$ for some $a, b \in R$, then $p \mid a$ or $p \mid b$.*

As noted before in the setting of ideals, the notion of irreducibility generalizes the idea of prime numbers having no divisors other than themselves, while the notion of primality is a generalization of the property $p \mid mn \implies p \mid m \vee p \mid n$ which characterizes prime numbers in $\mathbb{Z}$. Prime elements can be easily seen to be irreducible (if $p = ab$, then, since p divides itself, it must divide either a or b, being equivalent to it and forcing the other element to be a unit), but the converse is false in general. For instance, in $\mathbb{Z}[\sqrt{-5}]$, 3 may be seen to be irreducible, and it obviously divides $6 = (1 + \sqrt{-5})(1 - \sqrt{-5})$. However, it does not divide $1 + \sqrt{-5}$ nor $1 - \sqrt{-5}$ (something that may be seen by comparing norms) and thus it is not prime.

We could think at this point that it might be possible to choose the better behaved out of those two notions as our version of "building blocks" for a ring and develop a factorization theory from there. However, our example with $\mathbb{Z}[\sqrt{-5}]$ shows that neither is the proper one in some circumstances: 6 decomposes into irreducible numbers in two different ways, as $(1+\sqrt{-5})(1-\sqrt{-5})$ and $2 \cdot 3$, but no one of the numbers involved is prime, and we may show that there are no prime elements of $\mathbb{Z}[\sqrt{-5}]$ that are "up to the task", so to speak. In fact, the lack of prime numbers is what allows the existence of two different factorizations into irreducible elements. Thus, it is important to give a name to the "well-behaved" situation:

Definition 2.44 *An integral domain R is called a **unique factorization domain** (UFD) if every element $r \in R\setminus\{0\}$ may be written as a product of irreducible elements in a unique way up to equivalence. That is, for every $r \in R \setminus \{0\}$ there exist $u \in R^\times$ and $p_1, \dots, p_k \in R$ such that $r = up_1 \cdots p_k$, with all the p_j being irreducible; moreso, if $r = wq_1 \cdots q_\ell$ with $w \in R^\times$ and $q_1, \dots, q_\ell$ irreducible, then $k = \ell$ and there is some permutation $\sigma \in S_k$ such that p_i and $q_{\sigma(i)}$ are equivalent.*

We settle for irreducibles in the definition above because in the setting of UFDs it does not actually make a difference:

Proposition 2.45 *If R is a UFD, then every irreducible element is prime.*

We have met several examples of unique factorization domains already, as shown by:

Proposition 2.46 *Every principal ideal domain is a UFD.*

This includes every ring of polynomials $K[x]$ over a field, and algebraic integer rings such as $\mathbb{Z}[\mathrm{i}]$ and $\mathbb{Z}[\zeta_3]$. In these rings, the theory of divisibility can be developed in a similar way as what we know from $\mathbb{Z}$ (besides the additional care one must take with regards to units from $R^\times$) and an important tool in the context of algebraic integers is the field norm, because of its multiplicative properties. We have:

Proposition 2.47 *Let $K|\mathbb{Q}$ be a finite field extension, $N = N_{K|\mathbb{Q}}$ the associated field norm and $\mathcal{O}_K$ the corresponding ring of integers. We have the following properties:*

- *$N(x) \in \mathbb{Z}$ for every $x \in \mathcal{O}_K$,*
- *if $x \in R^\times$ (i.e., it is a unit) then $N(x) = \pm 1$, and similarly,*
- *if x and y are equivalent, then $N(x) = \pm N(y)$,*
- *$x \mid N(x)$; thus, if x is irreducible (hence prime), $N(x) = p^k$ for some prime integer p; more generally,*
- *$x \mid y \implies N(x) \mid N(y)$,*
- *$[\mathcal{O}_K : (x)] = |N(x)|$, where the left term is the index of the ideal (x) seen as an additive subgroup of $(\mathcal{O}_K, +)$.*

Thus, one may detect distinct irreducible factors of some $x \in \mathcal{O}_K$ by studying the prime factors of $N(x) \in \mathbb{Z}$, and units are automatically discarded in this fashion. While this does not eliminate all of the work involved (as $N(x)$ may be a power of the same prime p for non-equivalent irreducible elements of $\mathcal{O}_K$), for certain fields K the factorization problem reduces almost entirely to the equivalent problem in $\mathbb{Z}$.

2.3.2. Ideal factorization and Dedekind domains

What happens in the case where R is not a UFD? As it turns out, norms are still a very powerful tool for the task, and we will have something akin to unique factorization for any ring $\mathcal{O}_K$ when $[K : \mathbb{Q}] < \infty$. In particular, the existence of a factorization into irreducibles, even though it is not unique, can be proved via norms: either x is irreducible or $x = yz$ with y, z non-units. Since neither y nor z can be equivalent to x, then $|N(y)|$ and $|N(z)|$ are strictly smaller than $|N(x)|$, as they are nontrivial divisors of the latter. We can iterate this process on y and z, but only a finite number of times due to the well-ordering of $\mathbb{N}$; the end result is a factorization of x into irreducible terms. Thus, we need to see whether we can guarantee uniqueness, or if some generalization of this setting allows to “refine” two non-equivalent factorizations to result in the same decomposition.

Our solution will lean towards the latter option. Think of the previous example of failure of unique factorization, $\mathbb{Z}[\sqrt{-5}]$: we have that 6 has two essentially different factorizations into irreducibles, those being $2 \cdot 3$ and $(1+\sqrt{5})(1-\sqrt{5})$. One may think that, by introducing additional "numbers" to $\mathbb{Z}[\sqrt{-5}]$ (just like one adds the imaginary unit $\mathrm{i} = \sqrt{-1}$ to $\mathbb{R}$ to get a solution for the equation $x^2+1=0$), one may have a factorization such as:

$$2 = ab, \quad 3 = cd, \quad 1+\sqrt{5} = ac, \quad 1-\sqrt{5} = bd,$$

with a, b, c, d being some of those new "numbers", taking the role of prime or irreducible factors, such that $abcd$ is "the" factorization of 6. Kummer, who introduced this idea, called these extra elements "**ideal numbers**", and, as we shall see, the name is not a coincidence.

First, note that $\mathbb{Z}[\sqrt{-5}]$ is not a PID (as that would immediately imply that it is a UFD). The ideal (6) can be written as a product of two principal ideals in two ways coming from the previous factorizations; however, neither of the four ideals $(2), (3), (1+\sqrt{5}), (1-\sqrt{5})$ is actually a maximal ideal. Indeed, if we write $\mathfrak{a} = (2, 1+\sqrt{5})$, we see that both (2) and $(1+\sqrt{5})$ are contained in $\mathfrak{a}$; however, $3 \notin \mathfrak{a}$ and thus this ideal is nontrivial. Similarly, one may write $\mathfrak{b} = (2, 1-\sqrt{5})$, $\mathfrak{c} = (3, 1+\sqrt{5})$ and $\mathfrak{d} = (3, 1-\sqrt{5})$, and then it can be seen that:

$$(2) = \mathfrak{a} \cdot \mathfrak{b}, \quad (3) = \mathfrak{c} \cdot \mathfrak{d}, \quad (1+\sqrt{5}) = \mathfrak{a} \cdot \mathfrak{c}, \quad (1-\sqrt{5}) = \mathfrak{b} \cdot \mathfrak{d}.$$

Neither of these four ideals is principal, as that would imply the existence of irreducible nontrivial factors for numbers we already know are irreducibles. However, $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \mathfrak{d}$ are all maximal, as it can be verified that, e.g., $[\mathcal{O}_K : \mathfrak{a}] = 2$ so that there cannot be a nontrivial ideal that strictly contains $\mathfrak{a}$.

We previously mentioned that, for a general integral domain R, the notions of $m \mid n$, $n \in (m)$ and $(m) \supseteq (n)$ are all equivalent for $m, n \in R$. Thus, it makes sense to move our questions about divisibility to the domain of ideals, and no longer think directly on terms of elements of R unless it is strictly necessary. Thus, for two ideals $\mathfrak{a}$ and $\mathfrak{b}$, we write $\mathfrak{a} \mid \mathfrak{b}$ as a synonym for $\mathfrak{b} \subseteq \mathfrak{a}$ and we also write $\mathfrak{a} \mid b$ as a shorthand for $\mathfrak{a} \mid (b)$ (thus we can say that a given ideal divides some element of R). Again, we interpret principal ideals (a) as the set of multiples of a given number a; hence, we could interpret a general, non-principal ideal as the set of multiples of one of those "ideal numbers" R has been augmented with.

The following results lead us to our desired definition:

Proposition 2.48 *For any finite extension $K|\mathbb{Q}$, the ring of algebraic integers $\mathcal{O}_K$ is a quotient of some ring of polynomials $\mathbb{Z}[x_1, \ldots, x_d]$ and is thus Noetherian.*

Note that the fact that the extension $K|\mathbb{Q}$ is finite is essential to estabilish this result. For example, $\mathbb{E} = \mathcal{O}_{\mathbb{A}}$, the set of all algebraic integers, cannot be Noetherian, as it contains the infinite ascending chain $(2) \subset (\sqrt{2}) \subset (\sqrt[4]{2}) \subset (\sqrt[8]{2}) \subset \ldots$ which never stabilizes.

Proposition 2.49 *In a commutative Noetherian ring, every ideal $\mathfrak{a}$ can be written as a finite product of maximal ideals.*

Thus, we already have a method to factor ideals into a sort of equivalent of prime numbers. However, we lack uniqueness still; for this, we actually need primality, as this property of ideals is the one that actually ensures uniqueness of factorizations. Indeed, if an ideal $\mathfrak{a}$ can be written in two ways $\mathfrak{p}_1 \cdots \mathfrak{p}_r$ and $\mathfrak{q}_1 \cdots \mathfrak{q}_s$ as a product of prime ideals, each $\mathfrak{p}_i$ must divide one of the $\mathfrak{q}_j$ due to primality, and vice versa; under appropriate conditions and some work, this implies that each $\mathfrak{p}_i$ is one of the $\mathfrak{q}_j$ and that $r = s$.

Remember that an ideal $\mathfrak{p}$ of an integral domain R is prime if $R/\mathfrak{p}$ is also an integral domain, and maximal if this quotient is a field. Both situations are equivalent if this quotient is finite, and, since this is always the case[8] for rings of algebraic integers for a finite field extension $K|\mathbb{Q}$, we have the following:

Proposition 2.50 *If $K|\mathbb{Q}$ is a finite field extension, all nonzero prime ideals of $\mathcal{O}_K$ are also maximal (thus, the two notions are equivalent).*

In short, rings of algebraic integers for finite extensions are always in the following category:

Definition 2.51 *An integral domain R is a **Dedekind domain** if it is Noetherian, integrally closed and every nonzero prime ideal is maximal.*

This turns out to be exactly what we want for our purposes, as:

Theorem 2.52 *An integral domain R is a Dedekind domain if, and only if, every nonzero ideal $\mathfrak{a}$ has a unique factorization as a product of prime (equivalently, irreducible) ideals. This is, there exist prime ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_r$ such that $\mathfrak{a} = \mathfrak{p}_1 \cdots \mathfrak{p}_r$ and, if there are other prime ideals $\mathfrak{q}_1, \dots, \mathfrak{q}_s$ such that $\mathfrak{a} = \mathfrak{q}_1 \cdots \mathfrak{q}_s$, then $r = s$ and there exists some permutation $\sigma \in S_r$ such that $\mathfrak{p}_j = \mathfrak{q}_{\sigma(j)}, 1 \leq j \leq r$.*

To summarize, factorization of ideals is very well-behaved in our cases of interest:

Corollary 2.53 *For a finite field extension $K|\mathbb{Q}$, every ideal $\mathfrak{a}$ in $\mathcal{O}_K$ factors as a product of prime (equivalently, irreducible) ideals in a unique way.*

Thus, whenever we are analyzing a number ring that is not a UFD, we will use ideal factorization as our tool to answer questions of divisibility. These are often stated in terms of **valuations**, which, while for us will be little more than notational devices, are very important when delving deeper into number theory, field theory or algebraic geometry.

Definition 2.54 *Let K be a field. An (integer-valued[9]) **valuation** is a function $v\colon K \to \mathbb{Z} \cup \{\infty\}$ satisfying the following three properties:*

- $v(x) = \infty \iff x = 0$,
- $v(ab) = v(a) + v(b)$,
- $v(a+b) \geq \text{mín}(v(a), v(b))$.

[8] This is a consequence of the **integral basis lemma** from the next section, and of the fact that submodules of modules of finite rank over a PID are also of finite rank.

[9] In a general setting, it is useful to allow valuations whose values come from a general ordered group; however, we shall only need the $\mathbb{Z}$-valued version.

For example, the polynomial degree can be used to define a valuation in the field of rational fractions $\mathbb{Q}(x)$: it is clear that, for polynomials $p, q \in \mathbb{Q}[x]$, we have that $\deg(pq) = \deg(p) + \deg(q)$ and $\deg(p+q) \leq \text{máx}(\deg(p), \deg(q))$. Thus, defining $v(p/q) = \deg(q) - \deg(p)$ for nonzero polynomials p and q results in a function with the aforementioned three properties. Another example, which is the one we are interested in, is the following function defined in $\mathbb{Z}$ for a given prime number p:

$$v_p(n) := \begin{cases} \infty & \text{if } n = 0, \\ k & \text{if } p^k \mid n \wedge p^{k+1} \nmid n. \end{cases}$$

This extends to $\mathbb{Q}$ in the obvious way: $v_p(a/b) = v_p(a) - v_p(b)$, and can be verified to be a valuation in this field, called the **p-adic valuation**. The name is strongly related to the **group of p-adic integers** defined in the previous chapter, as the function $\delta_p(x, y) = p^{-v_p(x-y)}$ is exactly the restriction to $\mathbb{Z}$ of the p-adic metric (shift metric) we defined for $\mathbb{Z}_p$. Indeed, one may complete $\mathbb{Z}$ as a metric space using this p-adic valuation and obtain $\mathbb{Z}_p$ naturally.

The p-adic valuations satisfy the following equality:

$$(\forall x \in \mathbb{Q})\colon x = \prod_{p \in \mathbb{P}} p^{v_p(x)},$$

(where $\mathbb{P}$ represents the set of all integer primes) which, for integer values of x, equals their prime factorization. For general finite extensions, ideal factorization satisfies the same role; if we write $\mathbb{P}(K)$ for the set of all nontrivial prime ideals of $\mathcal{O}_K$ and let $\mathfrak{p} \in \mathbb{P}(K)$, the following defines a function that satisfies the three properties of a valuation in $\mathcal{O}_K$:

$$v_{\mathfrak{p}}(n) := \begin{cases} \infty & \text{if } n = 0, \\ k & \text{if } n \in \mathfrak{p}^k \setminus \mathfrak{p}^{k+1}, \end{cases}$$

and, since every element of K can be written as m/n for some $m, n \in \mathcal{O}_K$, we can extend the function $v_{\mathfrak{p}}$ to a valuation defined on all of K in the same way as in the integers. Thus, once again, the principal ideal (x) in $\mathcal{O}_K$ factors as:

$$(x) = \prod_{\mathfrak{p} \in \mathbb{P}(K)} \mathfrak{p}^{v_{\mathfrak{p}}(x)},$$

which will be our principal way to write ideal factorizations from now on.

As in the case of standard factorizations, the field norm is a very useful tool in this case. If we disregard signs, we may extend this to ideals themselves by reexamining their properties noted previously. Indeed, every ideal $\mathfrak{a}$ is an additive subgroup of $\mathcal{O}_K$, which is always of finite index when the extension $K|\mathbb{Q}$ is finite, and for principal ideals $\mathfrak{a} = (a)$ the value $[\mathcal{O}_K : \mathfrak{a}]$ is exactly $|N_{K|\mathbb{Q}}(a)|$. Thus, we define:

Definition 2.55 *The **norm** of an ideal $\mathfrak{a}$ of $\mathcal{O}_K$ in a finite field extension $K|\mathbb{Q}$ is its index as an additive subgroup of $\mathcal{O}_K$, i.e. $N_{K|\mathbb{Q}}(\mathfrak{a}) = [\mathcal{O}_K : \mathfrak{a}]$.*

Once again, $N_{K|\mathbb{Q}}(\mathfrak{a}) \in \mathfrak{a}$ (a fact we may write as $\mathfrak{a} \mid N_{K|\mathbb{Q}}(\mathfrak{a})$); as well, if $\mathfrak{a} \mid \mathfrak{b}$ (that is, $\mathfrak{a} \supseteq \mathfrak{b}$), then $N(\mathfrak{a}) \mid N(\mathfrak{b})$ and, similarly, $N_{K|\mathbb{Q}}(\mathfrak{a} \cdot \mathfrak{b}) = N_{K|\mathbb{Q}}(\mathfrak{a}) N_{K|\mathbb{Q}}(\mathfrak{b})$. Thus, we have the following result which shall be useful in later chapters:

Lemma 2.56 *Let $x \in \mathcal{O}_K$ for a finite field extension $K|\mathbb{Q}$. Then:*

$$N_{K|\mathbb{Q}}(x) = \pm \prod_{\mathfrak{p} \in \mathbb{P}(K)} N_{K|\mathbb{Q}}(\mathfrak{p})^{v_\mathfrak{p}(x)}.$$

Note that $N_{K|\mathbb{Q}}(\mathfrak{p}) = p^k$ for any prime ideal $\mathfrak{p}$ and some integer prime $p \in \mathbb{P}$, and hence $\mathfrak{p} \mid p$; that is, all prime ideals come from the factorization of integer primes in the ring $\mathcal{O}_K$. This allows us to classify the primes from $\mathbb{Z}$ (and the corresponding prime ideals) according to the behavior of their factorization in $\mathcal{O}_K$:

Definition 2.57 *Let $K|\mathbb{Q}$ be a finite field extension and $p \in \mathbb{P}$ an integer prime. We say that p is:*

- *an **inert prime** if the principal ideal (p) is prime in $\mathcal{O}_K$, i.e. the ideal factorization of (p) has exactly one term, which is (p) itself,*
- *a **split prime** if (p) factors as a product of distinct prime ideals, i.e. $(p) = \mathfrak{p}_1 \cdots \mathfrak{p}_r$ with $\mathfrak{p}_1, \dots, \mathfrak{p}_r \in \mathbb{P}(K)$ all distinct, and*
- *a **ramified prime** if there is some prime ideal $\mathfrak{p}$ such that $\mathfrak{p}^2 \mid (p)$, that is, the ideal factorization of (p) has some repeated factor.*

By abuse of terminology, we often use the terms inert, split and ramified to refer to the resulting prime ideals themselves.

This classification is important for our purposes because split and ramified primes (particularly the latter) introduce certain artifacts in our study of factorization that have implications of interest in the dynamical systems we shall define out of them. Determining which primes exhibit each of these behaviors (or a combination thereof) is something beyond the scope of this work; however, since we are particularly interested in ramified primes, it is worth noting the following result:

Theorem 2.58 *Let $K|\mathbb{Q}$ be a finite field extension and let $M\colon K \hookrightarrow \mathbb{R}^r \times \mathbb{C}^s$ be the Minkowski embedding defined above, with $r + 2s = d = [K : \mathbb{Q}]$. Let V be the least volume of a d-dimensional nondegenerate parallelogram whose vertices lie all in $M[\mathcal{O}_K] \subset \mathbb{R}^r \times \mathbb{C}^s$; the **discriminant** of K is the value $\Delta_K = \pm V^2$, which is always an integer*[10]*. Then, a prime p is ramified in $\mathcal{O}_K$ if, and only if, $p \mid \Delta_K$.*

The value Δ_K can be computed directly without appealing to the Minkowski embedding, and we shall indicate a method of doing so in a further section.

2.3.3. Dirichlet's unit theorem

We introduce now a small characterization of units in rings of algebraic integers, which we shall employ later on. Remember that the set $R^\times$ of units of a ring R is the collection of all elements that have multiplicative inverses; since $(xy)^{-1} = y^{-1}x^{-1}$, $R^\times$ is a group (which is abelian when R is commutative). We briefly discuss the structure of this group.

[10] The sign depends on the embeddings of the field K into $\mathbb{C}$; after we see an effective computation method for Δ_K below it will be clear that, since it involves determinants that may have imaginary terms, Δ_K can be negative.

Remember that a **root of unity** is a solution of the equation $x^n - 1 = 0$, that is, an element of the form $\zeta_n^k = e^{2\pi i k/n}$. Every n-th root of unity α that belongs to a given algebraic extension K of $\mathbb{Q}$ is a unit of $\mathcal{O}_K$, since $\alpha^{-1} = \alpha^{n-1}$ and α is naturally an algebraic integer. Thus, roots of unity are elements of finite order of the group $\mathcal{O}_K^\times$; conversely, if $u \in \mathcal{O}_K^\times$ is of finite order, it is necessarily a root of unity. This can be used to show that $\mathcal{O}_K^\times$ is a group of the form $U \times F$, where U is a group consisting of elements that are of infinite order[11] and F is a finite group (corresponding to $\mathcal{O}_K^\times \cap S^1$, with $S^1 = \{z \in \mathbb{C} \,:\, |z| = 1\}$).

To see that U may have nontrivial elements, think for example of the ring $\mathbb{Z}[\sqrt{2}] = \mathcal{O}_{\mathbb{Q}(\sqrt{2})}$. From the formula of the corresponding field norm, $N_{\mathbb{Q}(\sqrt{2})|\mathbb{Q}}(a + b\sqrt{2}) = a^2 - 2b^2$, we can see that, say, $\lambda = \sqrt{2} + 1$ (the case $a = b = 1$) is a unit with inverse $\sqrt{2} - 1$; however, its absolute value as an element from $\mathbb{C}$ is greater than 1, which implies that the numbers $(\sqrt{2} + 1)^n$ are all distinct for $n \in \mathbb{N}$, i.e. this is a unit of infinite order. It is not hard to convince oneself that every unit of $\mathbb{Z}[\sqrt{2}]^\times$ is of the form $(\pm 1 \pm \sqrt{2})^n$, and thus can be written as $\pm\lambda^n$ for $n \in \mathbb{Z}$, which implies that $\mathbb{Z}[\sqrt{2}]^\times \cong \mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z})$ (in our previous terminology, $U = \langle \lambda \rangle$ and $F = \{1, -1\}$ are the two roots of unity present in $\mathbb{R}$, as $\mathbb{Q}(\sqrt{2})$ is embedded into the reals).

As it turns out, there is a strong relationship between the possible embeddings of K into $\mathbb{C}$ and the units of $\mathcal{O}_K$. Remember that what we called a $\mathbb{Q}$-embedding is a ring homomorphism $\iota \colon K \hookrightarrow \mathbb{C}$ whose restriction to $\mathbb{Q}$ is the identity; we distinguished real embeddings and complex ones depending on whether the image $\mathrm{im}(\iota)$ was entirely contained in $\mathbb{R}$ or not, and mentioned that complex embeddings come in pairs (i.e. if ι is a complex embedding, then $\bar{\iota}(x) := \overline{\iota(x)}$ is another one). With this in mind, the relationship between these embeddings and the units of $\mathcal{O}_K$ is dictated by the following theorem:

Theorem 2.59 (Dirichlet's unit theorem) *Let $K|\mathbb{Q}$ be a finite field extension, and suppose K has exactly r real embeddings and s conjugate pairs of complex embeddings (such that $r + 2s = [K : \mathbb{Q}]$). Then U, the set of elements of infinite order of $\mathcal{O}_K^\times$, is isomorphic to the group $\mathbb{Z}^{r+s-1}$. More precisely, there exist $m = r + s - 1$ distinct units $\lambda_1, \ldots, \lambda_m \in \mathcal{O}_K^\times$, all multiplicatively independent (that is, $\lambda_1^{j_1} \cdots \lambda_m^{j_m} = 1$ implies that $j_1 = \cdots = j_m = 0$), such that every element of $\mathcal{O}_K^\times$ is of the form $\zeta\lambda_1^{j_1} \cdots \lambda_m^{j_m}$ for some root of unity ζ and $j_1, \ldots, j_n \in \mathbb{Z}$.*

One may compute the values of r and s without determining all of the embeddings. For instance, if we can find some α such that $K = \mathbb{Q}(\alpha)$, then the conjugates of α are exactly the n roots of $m_\alpha(x)$, and every embedding $K \hookrightarrow \mathbb{C}$ is entirely determined by which conjugate is chosen as the image of α; then, the embedding will be real if, and only if, this conjugate is real as well. Thus, r is the number of real roots of m_α, and consequently the remaining complex roots are exactly $2s = [K : \mathbb{Q}] - r$.

The generators of U are called the **fundamental units** of $\mathcal{O}_K$, and the set as a whole is called a **fundamental unit system**. In some casos, fundamental unit systems may be computed via geometric or algebraic considerations; for example, if K is a real field with $[K : \mathbb{Q}] = 2$, then the fundamental unit λ may be chosen as the element from $\mathcal{O}_K^\times$ with smallest absolute value greater than 1 (any of the two candidates may be chosen).

[11] The fundamental structure theorem for finitely generated $\mathbb{Z}$-modules implies, after some work, that U is then isomorphic to $\mathbb{Z}^r$ for some value of r.

2.3.4. Matrix representations for algebraic integers

This is, in a way, a continuation of the few notions of representation theory introduced in the previous chapter, and thus we once again we invite the reader to consult the book by Burrow [21] or other specialized sources; the computational aspects are also discussed by Jarvis [55]. After the discussion of the previous two sections, it is useful for computational purposes to find reasonable ways to embed rings of the form $\mathcal{O}_K$ (and, consequently, the whole field K) into the ring of matrices $\mathbb{M}_d(\mathbb{Q})$ for some appropriate $d \geq 1$. We start with a very useful definition that doubles as a lemma:

Lemma 2.60 (Integral basis lemma) *Every ring of algebraic integers $\mathcal{O}_K$ has an **integral basis**[12] B; that is, there exists a set of $d = [K : \mathbb{Q}]$ linearly independent elements $B = \{\omega_1, \ldots, \omega_d\} \subseteq \mathcal{O}_K$ such that $\mathcal{O}_K$ is the $\mathbb{Z}$-linear span of B, that is:*

$$\mathcal{O}_K = \{n_1\omega_1 + \ldots + n_d\omega_d \, : \, n_1, \ldots, n_d \in \mathbb{Z}\}.$$

Note that the set B is also a basis, in the linear algebra sense, of K seen as a $\mathbb{Q}$-vector space. We say that $\mathcal{O}_K$ is a d-dimensional (or **rank** d) $\mathbb{Z}$-**module**. In general, a module is essentially the same as a vector space, with the difference that scalar coefficients are allowed to belong to a ring instead of a field; we will not deal with many other examples of modules in this work, however.

Thus, we may identify $\mathcal{O}_K$ with $\mathbb{Z}^d$ via the usage of an integer basis. If we take some $\alpha \in K$, the function $M_\alpha(x) = \alpha x$ is linear (since multiplication is distributive) and thus has a representative matrix in the basis $B = \{\omega_1, \ldots, \omega_d\}$. Furthermore, if $\alpha \in \mathcal{O}_K$, then M_α must map any vector of $K \cong \mathbb{Q}^d$ with integer coordinates in the basis B to another vector with integer coordinates in this basis, which implies that the representative matrix[13] of M_α must necessarily have integer entries. In particular, if $\alpha \in \mathbb{Z}$, it is not hard to see that $M_\alpha = \alpha I_d$, a multiple of the identity matrix. It is not hard to verify that:

$$M_\alpha + M_\beta = M_{\alpha+\beta}, \quad M_\alpha \cdot M_\beta = M_{\alpha\cdot\beta},$$

and thus the mapping that sends α to the matrix M_α is a ring homomorphism $\varrho_B \colon \mathcal{O}_K \to \mathbb{M}_d(\mathbb{Z})$, that is, a **linear representation** of the ring $\mathcal{O}_K$ into the ring of matrices with integer entries. Furthermore, this representation is injective (that is, faithful), so it determines a subgroup of $\mathbb{M}_d(\mathbb{Z})$ that is isomorphic to $\mathcal{O}_K$.

It is also not hard to check that, if $\alpha \in \mathcal{O}_K^\times$, the matrix M_α has $M_{\alpha^{-1}}$ as its inverse, which is also a matrix with integer entries; thus, $M_\alpha \in \mathrm{GL}_d(\mathbb{Z})$ and hence $\det(M_\alpha) = \pm 1$. Hence, $\mathcal{O}_K^\times$ is a linear group under multiplication.

It is worth remembering, now, that the determinant function is a multiplicative homomorphism $\mathbb{M}_n(\mathbb{Q}) \to \mathbb{Q}$, this is, $\det(AB) = \det(A)\det(B)$. The field norm is also a multiplicative homomorphism that maps elements from $\mathcal{O}_K^\times$ to ± 1; from here, it is not hard to suspect the following:

[12] The definition of an integral basis also extends to the situation of an intermediate field $K \supset F \supset \mathbb{Q}$, in which elements of $\mathcal{O}_K$ are represented via linear combinations with coefficients in $\mathcal{O}_F$; however, in this more general situation integral bases may not exist (see Neukirch [81]). The proof of their existence relies, in particular, on the fact that $\mathbb{Z}$ is a PID.

[13] Below, we shall assume that we have a fixed integral basis B and thus we won't make distinctions between the linear transformation $x \mapsto \alpha x$ and its representative matrix, writing M_α for both.

Proposition 2.61 *Given an integral basis B for $\mathcal{O}_K$ and $\varrho_B\colon \mathcal{O}_K \to \mathbb{M}_d(\mathbb{Z})$ the corresponding linear representation, the following two equalities hold:*

$$\det(\varrho_B(\alpha)) = N_{K|\mathbb{Q}}(\alpha), \quad \operatorname{tr}(\varrho_B(\alpha)) = \operatorname{tr}_{K|\mathbb{Q}}(\alpha).$$

Note that the same equalities hold for the obvious extension of ϱ_B to the whole of K. It is also easy to note that, for any polynomial $p \in \mathbb{Q}[x]$, $p(M_\alpha) = M_{p(\alpha)}$; thus, since any $\alpha \in K$ is algebraic over $\mathbb{Q}$, its minimal polynomial $m_\alpha \in \mathbb{Q}[x]$ satisfies the equality $m_\alpha(M_\alpha) = 0_d$, the zero matrix. Hence, we have the following as a consequence of the Cayley-Hamilton theorem:

Proposition 2.62 *Given $K|\mathbb{Q}$ a finite field extension, B an integral basis for $\mathcal{O}_K$, and ϱ_B the corresponding linear representation, let $\alpha \in K$ and $M = M_\alpha = \varrho_B(\alpha)$ the associated matrix. Then, the minimal polynomial of the matrix M (that is, the least degree polynomial $p \in \mathbb{Q}[x]$ such that $p(M) = 0$) is $m_\alpha(x)$, the minimal polynomial of α in the field extension sense. Furthermore, the characteristic polynomial of M is a power of m_α; more specifically, it equals:*

$$\chi_M(x) = m_\alpha(x)^{[K:\mathbb{Q}(\alpha)]}.$$

The previous two results give us powerful tools to compute the norm, trace and minimal polynomial of algebraic integers upon finding an appropriate integral basis. For example, $\{1, \sqrt[3]{2}, \sqrt[3]{4}\}$ is an integral basis for $\mathbb{Q}(\sqrt[3]{2})$, upon which the representation homomorphism ϱ_B is given by:

$$\varrho_B(\sqrt[3]{2}) = \begin{bmatrix} 0 & 0 & 2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \implies \varrho_B(a + b\sqrt[3]{2} + c\sqrt[3]{4}) = \begin{bmatrix} a & 2c & 2b \\ b & a & 2c \\ c & b & a \end{bmatrix},$$

and thus the norm and trace are now easy to compute:

$$N_{\mathbb{Q}(\sqrt[3]{2})|\mathbb{Q}}(a + b\sqrt[3]{2} + c\sqrt[3]{4}) = \begin{vmatrix} a & 2c & 2b \\ b & a & 2c \\ c & b & a \end{vmatrix} = a^3 + 2b^3 + 4c^3 - 6abc,$$
$$\operatorname{tr}_{\mathbb{Q}(\sqrt[3]{2})|\mathbb{Q}}(a + b\sqrt[3]{2} + c\sqrt[3]{4}) = 3a.$$

To compute the norm directly from the definition, we would have required to find explicitly the three embeddings of $\mathbb{Q}(\sqrt[3]{2})$ into $\mathbb{C}$ or find a Galois extension K of $\mathbb{Q}$ that contains $\mathbb{Q}(\sqrt[3]{2})$; a similar situation applies to the trace. Furthermore, if we wanted to compute the minimal polynomial of, say, $1 + \sqrt[3]{4}$, we could do so as follows:

$$M = \varrho_B(1 + \sqrt[3]{4}) = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 1 & 0 & 1 \end{bmatrix},$$
$$\implies \chi_M(\lambda) = \det(\lambda I_3 - M) = \begin{vmatrix} \lambda - 1 & -2 & 0 \\ 0 & \lambda - 1 & -2 \\ -1 & 0 & \lambda - 1 \end{vmatrix}$$
$$= (\lambda - 1)^3 - 4,$$
$$\therefore m_{1+\sqrt[3]{4}}(x) = x^3 - 3x^2 + 3x - 5.$$

Finding an appropriate integral basis may be hard. It is a known fact that any finite field extension $K|\mathbb{Q}$ may be taken to be of the form $K = \mathbb{Q}(\alpha)$ for some appropriate $\alpha \in K$, and thus $B_\alpha = \{1, \alpha, \alpha^2, \ldots, \alpha^{d-1}\}$ is a basis of K as a $\mathbb{Q}$-vector space, where $d = [K : \mathbb{Q}]$. However, this α need not be an algebraic integer, or, if it is, the $\mathbb{Z}$-linear span of B_α may be a strict submodule of $\mathcal{O}_K$ (see, e.g. $\mathbb{Q}(\sqrt{5})$, where $\sqrt{5}$ is an algebraic integer, but $\mathbb{Z}[\sqrt{5}]$, the linear span of $\{1, \sqrt{5}\}$, does not contain the golden ratio $\varphi = \frac{1}{2}(1 + \sqrt{5})$ despite it also being an algebraic integer, as a root of $x^2 - x - 1$). Furthermore, it is known that there exists some fields K such that $\mathcal{O}_K$ has no integral basis of this form. We note that such bases, if they do exist, are relatively easier to handle, as if $m_\alpha(x) = x^d + a_{d-1}x^{d-1} + \cdots + a_0$ is the minimal polynomial of α, then, in the basis $B = B_\alpha$, we have:

$$M_\alpha = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{d-1} \end{bmatrix} \implies \varrho_B\left(\sum_{k=0}^{d-1} b_k \alpha^k\right) = \sum_{k=0}^{d-1} b_k M_\alpha^k,$$

which is often friendly to computations.

An additional application of integral bases is the effective computation of the discriminant Δ_K defined previously in connection with ramified primes. We have the following result:

Lemma 2.63 *Let $K|\mathbb{Q}$ be a finite field extension of degree $d = [K : \mathbb{Q}]$, $\{\omega_1, \ldots, \omega_d\}$ be an integral basis of $\mathcal{O}_K$, and $\{\iota_1, \ldots, \iota_d\}$ be the set of all (real and complex) $\mathbb{Q}$-embeddings of K. Then, the discriminant may be computed as any of the two following determinants:*

$$\Delta_K = \begin{vmatrix} \iota_1(\omega_1) & \iota_1(\omega_2) & \cdots & \iota_1(\omega_d) \\ \iota_2(\omega_1) & \iota_2(\omega_2) & \cdots & \iota_2(\omega_d) \\ \vdots & \vdots & \ddots & \vdots \\ \iota_d(\omega_1) & \iota_d(\omega_2) & \cdots & \iota_d(\omega_d) \end{vmatrix}^2 = \begin{vmatrix} \operatorname{tr}_{K|\mathbb{Q}}(\omega_1\omega_1) & \operatorname{tr}_{K|\mathbb{Q}}(\omega_1\omega_2) & \cdots & \operatorname{tr}_{K|\mathbb{Q}}(\omega_1\omega_d) \\ \operatorname{tr}_{K|\mathbb{Q}}(\omega_2\omega_1) & \operatorname{tr}_{K|\mathbb{Q}}(\omega_2\omega_2) & \cdots & \operatorname{tr}_{K|\mathbb{Q}}(\omega_2\omega_d) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{tr}_{K|\mathbb{Q}}(\omega_d\omega_1) & \operatorname{tr}_{K|\mathbb{Q}}(\omega_d\omega_2) & \cdots & \operatorname{tr}_{K|\mathbb{Q}}(\omega_d\omega_d) \end{vmatrix}.$$

This also gives a theoretical (albeit not very practical) method to find an integral basis, by looking, among all linearly independent d-tuples of algebraic integers of K, the ones that minimize the value of this determinant.

2.3.5. Quadratic fields and number rings

Now, we briefly summarize the properties and classifications held by fields K of index 2 and their associated rings of algebraic integers; this is based on the discussion in the book by Jarvis [55]. As we shall see, these fields are always of the form $\mathbb{Q}(\sqrt{d})$, where $d \in \mathbb{Z}$ is a square-free number (i.e., for any $p \in \mathbb{P}$, $p^2 \nmid d$), and their properties can be linked directly to the value of d.

Definition 2.64 *Any field K with $[K : \mathbb{Q}] = 2$ is called a* ***quadratic field****. They are always of the form $\mathbb{Q}(\sqrt{d})$ with d a square-free integer. When $d > 0$, we speak of a* ***real quadratic field****, and an* ***imaginary quadratic field*** *otherwise; this is because $\mathbb{Q}(\sqrt{d}) \subseteq \mathbb{R}$ has two real embeddings when $d > 0$ and none when $d < 0$.*

Both real and imaginary quadratic fields $\mathbb{Q}(\sqrt{d})$ have only one nontrivial Galois automorphism, which is conjugation $a + b\sqrt{d} \mapsto a - b\sqrt{d}$. When $d < 0$, this coincides with complex conjugation, and thus $\mathbb{Q}(\sqrt{d})$ has a conjugate pair of complex embeddings ($s = 1, r = 0$). For real quadratic fields, there are two real embeddings (i.e. $r = 2, s = 0$). In both cases, $\mathbb{Q}(\sqrt{d})$ is trivially a Galois extension of $\mathbb{Q}$ with $|\mathrm{Gal}(\mathbb{Q}(\sqrt{d})|\mathbb{Q})| = 2$.

We mentioned previously that sometimes $\mathcal{O}_{\mathbb{Q}(\sqrt{d})}$ is not $\mathbb{Z}[\sqrt{d}]$. This actually depends on the value of d, and is easily summarized by the following:

Proposition 2.65 *If $d \neq 4k + 1$ for some $k \in \mathbb{Z}$, then the ring of algebraic integers of $\mathbb{Q}(\sqrt{d})$ is $\mathbb{Z}[\sqrt{d}]$, and thus $\{1, \sqrt{d}\}$ is an integral basis for $\mathbb{Q}(\sqrt{d})$. When $d = 4k+1$, we have $\mathcal{O}_{\mathbb{Q}(\sqrt{d})} = \mathbb{Z}[\delta]$ with $\delta = \frac{1}{2}(1 + \sqrt{d})$, which implies that $\{1, \delta\}$ is an integral basis.*

For example, $\mathcal{O}_{\mathbb{Q}(\sqrt{5})} = \mathbb{Z}[\varphi]$, with φ the golden ratio, and $\mathbb{Z}[\sqrt{5}]$ is a strict subring (a non-maximal order).

Dirichlet's theorem gives us the following characterization of units in quadratic fields:

Corollary 2.66 *Imaginary quadratic fields have always a finite number of units, which correspond to roots of unity: four for $d = -1$, six for $d = -3$ and two for every other negative value of d. In contrast, real quadratic fields have infinitely many units, and thus $\mathcal{O}^{\times}_{\mathbb{Q}(\sqrt{d})} \cong \mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z})$, and all of the units are of the form $\pm(a + b\sqrt{d})^n, n \in \mathbb{Z}$ where (a, b) is the smallest positive solution of the associated **Pell's equation**[14] $a^2 - db^2 = 1$.*

The ring $\mathcal{O}_{\mathbb{Q}(\mathrm{i})} = \mathbb{Z}[\mathrm{i}]$ is often called the ring of **Gaussian integers**. Similarly, $\mathcal{O}_{\mathbb{Q}(\sqrt{-3})} = \mathbb{Z}[\zeta_3]$ is known as the ring of Eisenstein integers. Both are PIDs and therefore unique factorization domains; furthermore, they are **Euclidean domains**, meaning that they have a division algorithm just like the one of the integers. To be more precise:

Definition 2.67 *We say that an integral domain R is **Euclidean** if there exists some function $\varphi\colon \mathbb{R} \setminus \{0\} \to \mathbb{N}$ such that, for every $a, b \in R$, there exist $q, r \in R$ for which $a = bq + r$ and either $r = 0$ or $\varphi(r) < \varphi(b)$. Such a function is called an **Euclidean function**.*

That is, the function φ is the one that gives a notion of the remainder r being "smaller" than b in some sense. In the usual division algorithm in $\mathbb{Z}$, we use the absolute value φ as our Euclidean function; for polynomials over a field K, the degree is the most obvious candidate. Since the existence of a division algorithm allows us to compute greatest common divisors using Euclid's algorithm, this is a property that is strictly stronger than being a PID. It is known that:

Theorem 2.68 *An imaginary quadratic field $\mathbb{Q}(\sqrt{d}), d < 0$ has an Euclidean ring of algebraic integers only when $d = -1, -2, -3, -7, -11$. This ring is a non-Euclidean principal ideal domain only when $d = -19, -43, -67, -163$.*

It is conjectured (but not known) that there are infinitely many real quadratic fields that are PIDs; the status of Euclidean domain among them is, thus, also unknown.

[14] This equation can be solved via continued fraction techniques.

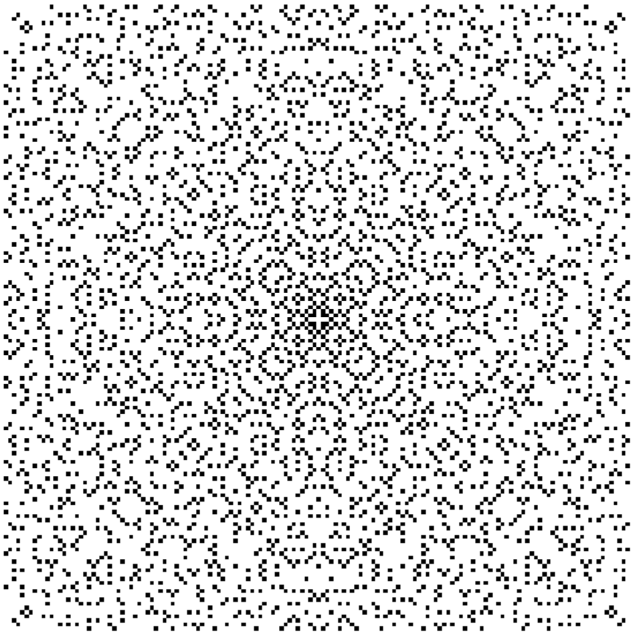


Figure 2.1: Primes in the ring of Gaussian integers $\mathbb{Z}[i]$; this is an example of the ring of integers of a quadratic field.

We conclude with some brief notes on the classification of prime numbers in some quadratic field K. First, using the integral bases we noted above, we obtain the following:

Proposition 2.69 *The discriminant Δ_K of $K = \mathbb{Q}(\sqrt{d})$ is given by:*

$$\Delta_K = \begin{cases} d & \textit{if } d = 4k+1,\ k \in \mathbb{Z}, \\ 4d & \textit{otherwise.} \end{cases}$$

Given our characterization of ramification above, we conclude then that:

Corollary 2.70 *A rational prime $p \in \mathbb{P}$ ramifies in $\mathcal{O}_{\mathbb{Q}(\sqrt{d})}$ if it divides d. If $d \not\equiv 1 \pmod 4$, then 2 is also a ramified prime.*

For the remaining primes, there is also a verifiable condition to check whether they split or not:

Proposition 2.71 *Let p be a rational prime that does not ramify in $\mathbb{Q}(\sqrt{d})$. Then, p splits if, and only if, there exists some $b \in \mathbb{Z}$ such that $d - b^2$ is divisible*[15] *by p.*

This is a consequence of the fact that $d - b^2$ factors as $(\sqrt{d} + b)(\sqrt{d} - b)$ in $\mathcal{O}_{\mathbb{Q}(\sqrt{d})}$, and the primality property for ideals. Using techniques such as the quadratic reciprocity theorem, it is then possible to classify every prime in the ring of integers of a quadratic field as inert ($(p) = \mathfrak{p}$), split ($(p) = \mathfrak{p}\bar{\mathfrak{p}}$) or ramified ($(p) = \mathfrak{p}^2$).

[15] We say that d is a **quadratic residue** modulo p, that is, there exists a solution b in $\mathbb{Z}/p\mathbb{Z}$ of the equation $x^2 \equiv d \pmod p$.

2.4. Basic notions of algebraic geometry

As a complement to the previous section, we briefly introduce some basic notions of algebraic geometry, used in connection to the study of norms in integer rings. The standard references in this subject are the books by Hartshorne [51] (albeit to a level of detail unneeded in the current work) and Shafarevich [93], and the book by Lang [69] has an overview of the more classical aspects of the theory. Gathmann [45] and Kendig [60] provide a simpler introduction to the basic underlying concepts we shall need, along with Vaisencher [95] for the case of plane curves..

Algebraic geometry deals with the set of solutions of algebraic equations in a field, such as, e.g. $S^1 = \{(x, y) \in \mathbb{R}^2 \,:\, x^2 + y^2 = 1\}$, which corresponds to a plane curve (specifically, a circle). Similarly, equations with more variables may be thought of as defining surfaces or manifolds in appropiate ambient spaces. However, these sets of solutions often exhibit phenomena not usually taken into account in the standard theory of curves, e.g. self-intersections or multiplicities. For instance, the **Descartes folium** is a curve, given by the equation $x^3 + y^3 = 3xy$, that intersects itself, while the equation $(x + y - 1)^2 = 0$ describes the same points as the equation $x + y = 1$, which corresponds to a straight line; however, each point on the first equation is “accounted for” twice, and we may think of this equation as representing two straight lines, one on top of the other. This is, in a way, a distinction of algebraic nature.

Another problem that may arise is that a given set V may need more than one equation to describe it, or that there may be several different equations appearing naturally that result in the same set V. Thus, in the algebraic-geometric viewpoint, we identify each curve (or surface, or etc.) with the corresponding family of (systems of) equations defining it, a viewpoint which naturally leads to ideals: if $p(x, y) = 0$ along every point of a curve, then, for any polynomial $q(x, y)$, we will have $q(x, y)p(x, y) = 0$ as well. Formally, we proceed as follows:

Definition 2.72 *Let K be a field and $K[x_1, \dots, x_d]$ be the corresponding ring of polynomials in d variables. Given an ideal $\mathfrak{r} \subseteq K[x_1, \dots, x_d]$ in the latter ring, the **affine variety** (or **zero locus**, or **algebraic set**; in the particular case of $d = 2$ we often speak of an **algebraic curve**) defined by $\mathfrak{r}$ corresponds to the following subset of K^d:*

$$\mathsf{V}(\mathfrak{r}) := \{(u_1, \dots, u_d) \,:\, (\forall p(x_1, \dots, x_d) \in \mathfrak{r})\colon p(u_1, \dots, u_d) = 0\}.$$

*Reciprocally, given any set $X \subseteq K^d$, we define the **ideal** of X as the following set of polynomials:*

$$\mathfrak{I}(V) := \{p \in K[x_1, \dots, x_d] \,:\, (\forall (u_1, \dots, u_d) \in X)\colon p(u_1, \dots, u_d) = 0\}.$$

If $\mathfrak{r} = (p_1, \dots, p_k)$ is the ideal generated by the polynomials $p_1, \dots, p_k$, we may use the notation $\mathsf{V}(p_1, \dots, p_k)$ instead of $\mathsf{V}(\mathfrak{r})$ for the corresponding variety.

Affine varieties represent our idea of curves, surfaces, etc., described by algebraic equations; for instance, the unit circle $S^1 \subseteq \mathbb{R}^2$ corresponds to the affine variety $\mathsf{V}(x^2 + y^2 - 1)$ in the reals, while the straight line in $\mathbb{R}^3$ that passes through $(0, 0, 0)$ and $(1, 1, 1)$ corresponds to $\mathsf{V}(x - y, x - z)$, which can be seen as the intersection of the two hyperplanes $\mathsf{V}(x - y)$ and $\mathsf{V}(x - z)$.

The sets $\mathsf{V}(\mathfrak{r})$ and $\mathfrak{I}(X)$ are related by the following properties:

Proposition 2.73 *Given $\mathfrak{r}, \mathfrak{s}$ ideals in $K[x_1, \ldots, x_d]$ and $X, Y \subseteq K^d$, we have the following:*

(i) $X \subseteq Y \implies \mathfrak{I}(Y) \subseteq \mathfrak{I}(X)$.

(ii) $\mathfrak{r} \subseteq \mathfrak{s} \implies \mathsf{V}(\mathfrak{s}) \subseteq \mathsf{V}(\mathfrak{r})$.

(iii) $X \subseteq \mathsf{V}(\mathfrak{I}(X))$, *with equality when X is an affine variety.*

(iv) $\mathfrak{r} \subseteq \mathfrak{I}(\mathsf{V}(\mathfrak{r}))$.

To clarify the latter relationship, we need to introduce an additional concept:

Definition 2.74 *Let R be any integral domain and $\mathfrak{a}$ an ideal in R. The* ***radical*** *of $\mathfrak{a}$ is the following ideal of R:*

$$\sqrt{\mathfrak{a}} := \{r \in R \;:\; (\exists n \in \mathbb{N})\colon r^n \in \mathfrak{a}\}.$$

Note that if, for a set X we have that $p(x)^n = 0$ for all $x \in X$, then $p(x) = 0$ for all $x \in X$. Thus, if $p(x)^n \in \mathfrak{I}(X)$, then $p(x) \in \mathfrak{I}(X)$, i.e. $\sqrt{\mathfrak{I}(X)} = \mathfrak{I}(X)$. An ideal with this property is called a **semiprime ideal** or **radical ideal**. We have, thus, that:

Theorem 2.75 *If K is an algebraically closed field and $\mathfrak{r}$ an ideal of $K[x_1, \ldots, x_d]$, then $\mathfrak{I}(\mathsf{V}(\mathfrak{r})) = \sqrt{\mathfrak{r}}$; thus, equality holds if and only if $\mathfrak{r}$ is a radical ideal. We have a 1-1 correspondence between affine varieties in K^d and radical ideals in $K[x_1, \ldots, x_d]$, given by the maps $\mathfrak{I}$ and V.*

The previous theorem is a form of Hilbert's *Nullstellensatz*, one of the central theorems in algebraic variety. Another form, which is the one that is more commonly found, is the one that ensures that algebraic varieties in algebraically closed fields are nonempty. Indeed, we say that a **zero of an ideal** $\mathfrak{r} \subseteq K[x_1, \ldots, x_d]$ is a point $(u_1, \ldots, u_d)$ such that $p(u_1, \ldots, u_d) = 0$ for every $p \in \mathfrak{r}$; then, Hilbert's theorem says:

Theorem 2.76 (Hilbert's *Nullstellensatz*) *Let K be an algebraically closed field. If $\mathfrak{r}$ is a nontrivial ideal of $K[x_1, \ldots, x_d]$, then $\mathfrak{r}$ has a zero in K^d.*

Note that for $d = 1$ this is a restatement of the fact that K is algebraically closed, and thus this hypothesis cannot be weakened.

The theory of algebraic sets and varieties is much more developed for the situation where K is algebraically closed, as tools such as the previous theorem and the guarantee that every one-variable polynomial decomposes as a product of linear factors are routinely employed in this context. Thus, most of the facts below will apply to the space $\mathbb{C}^d$, and we shall keep in mind that they do not always translate directly to our context (e.g. it will not be obvious how to study the set of points with integer coordinates of an algebraic variety) and we might need some additional work.

We now introduce some basic properties of affine varieties in relation to ideals. Firstly, we have the following:

Proposition 2.77 *Let V_1 and V_2 be affine varieties. Then $V_1 \cap V_2$ and $V_1 \cup V_2$ are affine* *varieties. In particular, if $V_i = \mathsf{V}(\mathfrak{r}_i)$, we have that:*

$$\begin{aligned} V_1 \cup V_2 &= \mathsf{V}(\mathfrak{r}_1 \cdot \mathfrak{r}_2) = \mathsf{V}(\mathfrak{r}_1 \cap \mathfrak{r}_2), \\ V_1 \cap V_2 &= \mathsf{V}(\mathfrak{r}_1 + \mathfrak{r}_2). \end{aligned}$$

The latter relationship may be extended to any infinite collection of ideals, by noting that the set $\bigoplus_{i \in I} \mathfrak{r}_i$ (the set of all finite sums of terms from $\bigcup_{i \in I} \mathfrak{r}_i$) is an ideal as well, that contains every $\mathfrak{r}_i$. Thus, we have:

$$\bigcap_{i \in I} V_i = \mathsf{V}\left(\bigoplus_{i \in I} \mathfrak{r}_i\right),$$

and hence the arbitrary intersection of affine varieties is also an affine variety.

The previous result may be restated as: the set of all algebraic varieties over the space K^d is the collection of closed sets of a topology on K^d. We call this topology the **Zariski topology**; it has the advantage that we may use the language of topological spaces when describing variety-related phenomena, such as connectedness.

As noted before, there are some varieties made of "pieces" or "components" which are themselves varieties. For instance, the variety $x^2 - y^2 = 0$ in $\mathbb{R}^2$ consists of two straight lines $x \pm y = 0$. We thus distinguish between varieties that have such a decomposition from those which don't:

Definition 2.78 *An affine variety V is **reducible** if it is the union of two nonempty affine varieties V_1, V_2, both strictly smaller than V; otherwise, we say that V is **irreducible**. A **component** of V is a maximal irreducible affine subvariety $V' \subseteq V$.*

Irreducibility, which may be thought of as a purely geometric notion, is pretty much algebraic as well. Indeed, it can be shown that:

Definition 2.79 *Let $\mathfrak{r} \subseteq K[x_1, \ldots, x_d]$ be a radical ideal in an algebraically closed field. Then, the affine variety $\mathsf{V}(\mathfrak{r})$ is irreducible if, and only if, the ideal $\mathfrak{r}$ is prime.*

In the previous example given by the equation $x^2 - y^2 = 0$, we see that the polynomial $x^2 - y^2$ is not irreducible, and factors as $(x - y)(x + y)$, each of the factors corresponding to one of the two lines we identified as irreducible components. In contrast, the hyperbola defined by the equation $x^2 - y^2 = 1$ is irreducible and its only component is itself, as the polynomial $x^2 - y^2 - 1$ is irreducible, and the corresponding ideal prime. Intuitively, if an ideal $\mathfrak{r}$ in $K[x_1, \ldots, x_d]$ factors as a product of prime ideals $\mathfrak{p}_1 \cdots \mathfrak{p}_k$, the irreducible components of $\mathsf{V}(\mathfrak{r})$ should be $\mathsf{V}(\mathfrak{p}_j)$, $1 \leq j \leq k$. We have:

Theorem 2.80 *Let V be an affine variety in K^d, with K an algebraically closed field. Then there exist finitely many irreducible varieties $V_1, \ldots, V_k$, with $V_i \not\subseteq V_j$ for $i \neq j$, such that $V = V_1 \cup \cdots \cup V_k$.*

Since every irreducible component $V' \subseteq V$ is equal to the union of the affine varieties $V' \cap V_j$, then it must be equal to one of the V_j, as otherwise it contradicts either the irreducibility or the maximality of V. Thus, this decomposition into irreducible components is unique up to reordering, and $V_1, \ldots, V_k$ are the only irreducible components of V.

It is common in algebraic geometry to work in projective space $K\mathbf{P}^d$ instead of usual affine space K^d, as this makes the statement of certain theorems easier. Projective space is obtained by endowing K^d with additional "points at infinity" such that, for instance, two parallel lines in K^d intersect in one of these new points. Formally, this corresponds to a quotient $(K^{d+1} \setminus \{\mathbf{0}\})/\sim$, where $\sim$ is the equivalence relation $\boldsymbol{x} \sim \boldsymbol{y}$ iff $\boldsymbol{x} = \lambda\boldsymbol{y}$ for some nonzero $\lambda \in K$. When K (and thus K^d) has a topological space structure, we give the associated quotient topology to $K\mathbf{P}^d$.

Points in $K\mathbf{P}^d$ may be associated a set of $d+1$ coordinates $[u_1, \dots, u_{d+1}]$, not all of them 0, with the understanding that $[\lambda u_1, \dots, \lambda u_{d+1}]$ represents the same point for any $\lambda \neq 0$. The set of points with coordinates $[u_1, \dots, u_d, 1]$, in particular (or any other set with one coordinate fixed as a nonzero value) is in a 1-1 bijection with K^d, while the additional elements of $K\mathbf{P}^d$ may be thought of as points at infinity (which by themselves have the same structure as $K\mathbf{P}^{d-1}$).

Any affine variety in K^d extends to $K\mathbf{P}^d$ in a natural way as a quotient of a $(d+1)$-dimensional affine variety under the equivalence relation $\sim$; for example, the extension of the parabola $y - x^2 = 0$ to projective space is the set of all points $[x, y, t] \in \mathbb{R}\mathbf{P}^2$ that satisfy the equation $ty - x^2 = 0$, which are exactly the same points $[x, y, 1]$ from the original parabola together with an additional point $[0, 1, 0]$ which makes the resulting curve homeomorphic to a circle. Note that the 3-dimensional affine variety $\mathsf{V}(ty - x^2)$ has the property that, whenever $(x, y, t) \in \mathsf{V}(ty - x^2)$, we have $(\lambda x, \lambda y, \lambda t) \in \mathsf{V}(ty - x^2)$ as well, so as to make the quotient well-defined; a similar procedure can be made for any affine variety given by some ideal $\mathfrak{r}$. We call the quotient of a $(d+1)$-dimensional affine variety from K^{d+1} under $\sim$ a d-dimensional **projective variety**.

Now that we have the basic terminology of algebraic geometry, we introduce a few theorems of common usage that we might encounter later on, including some properties of conic sections we will find useful later on. First, we note that, while we are appealing to an intuitive definition of **dimension** $\dim(V)$ of a variety V, in the case of varieties it may formally be defined as the longest possible chain of inclusions $V \supset V' \supset V'' \supset \cdots \supset V^{(n)}$ where each $V^{(i)}$ is irreducible and the inclusions are strict. This coincides with our intuitive notion of dimension whenever the variety looks locally like a d-dimensional space in a topological sense. Thus, it makes sense to define the **codimension** of V as $\operatorname{cod}(V) := d - \dim(V)$.

Theorem 2.81 (Dimension theorem) *For two affine (or projective) varieties V_1 and V_2, we have:*

$$\operatorname{cod}(V_1 \cap V_2) \leq \operatorname{cod}(V_1) + \operatorname{cod}(V_2).$$

Note that for affine varieties $V_1 \cap V_2$ might be empty, so we define $\operatorname{cod}(\varnothing) = -1$.

The following is an important result about intersections of varieties with codimension 1. We only use a very weak version of this theorem, but for completeness we state it in full:

Theorem 2.82 (Bézout) *Let $p_1, \dots, p_d \in K[x_1, \dots, x_d]$ be n polynomials over an algebraically closed field K, and $\mathsf{V}(p_1), \dots, \mathsf{V}(p_d)$ the associated projective varieties, all of them of codimension 1. Then, either:*

- $\bigcap_{k=1}^{d} \mathsf{V}(p_k)$ *is infinite, or*

- *the number of intersections, counting multiplicities, is* $\prod_{k=1}^{d} \deg(p_k)$.

Additional conditions may be introduced to ensure that the finite case happens; however, they are too technical for the scope of this work. We present a simple corollary of this result, which may be proved by more elementary methods and is of interest to us; remember that a **conic section** is a plane curve whose equation is of the form:

$$ax^2 + bxy + cy^2 + dx + ey + f = 0.$$

Corollary 2.83 (Five points determine a conic) *Let V_1 and V_2 be two irreducible conic sections (that is, neither is the union of two straight lines). If $|V_1 \cap V_2| \geq 5$, then $V_1 = V_2$.*

In particular, any two irreducible, distinct conic sections intersect in exactly four points, counting multiplicities (that is, points where both curves are tangent), complex intersections and points at infinity.

Finally, to close this section, we introduce an important result on mappings between algebraic varieties, in the version that is useful for us:

Theorem 2.84 (Ax–Grothendieck) *Let $p = (p_1, \ldots, p_d) \colon K^d \to K^d$ be a function of K^d to itself (with K algebraically closed) such that every coordinate p_j is a polynomial function on the variables $x_1, \ldots, x_d$, and suppose $p[V] \subseteq V$ for an affine (or projective) variety V. Then, if p is injective, it must be surjective as well.*

The original version of this theorem, proved by both James Ax and Alexander Grothendieck independently, was stated for $V = \mathbb{C}^d$ and used model theory in an interesting way to derive a contradiction. Our main interest is in the affine subcase: if V is a variety and $M \cdot V + \boldsymbol{x} \subseteq V$ for some invertible matrix M and $\boldsymbol{x} \in K^d$, then $M \cdot V + \boldsymbol{x} = V$.

Chapter 3

Multidimensional symbolic dynamics, aperiodic order and affine topics

In this chapter we introduce the main concepts of symbolic dynamics we will need for this work. The classic references for one-dimensional symbolic dynamics are the books by Lind and Marcus [71] and Kitchens [62], together with Kůrka [64] for a description of certain subshifts with interesting dynamical behavior. One-dimensional substitutions are studied in several books, including Lothaire [72, 73] and Pytheas Fogg [38]. Finally, we direct the reader to Ceccherini-Silberstein and Coornaert's book [24], as, while it is not focused on symbolic dynamics proper, it introduces most of the new terminology that distinguishes the multidimensional situation from the more studied one-dimensional case.

3.1. Shift spaces

The symbol $\mathcal{A}$ will be reserved for a finite set (**alphabet**) and its elements will be referred to as **symbols**. Since we deal with multidimensional subshifts, we shall reserve the letter d to refer to the dimension of the space, and use vector notation $\boldsymbol{k} = (k_1, \ldots, k_d)$ to refer to specific elements of $\mathbb{Z}^d$. The letter $\boldsymbol{s}$ and corresponding tuple $(s_1, \ldots, s_d)$, in particular, will be reserved for a specific "size" number associated to a specific substitution.

In what follows, we shall deal with spaces of functions or **configurations**[1] of symbols $x\colon \mathbb{Z}^d \to \mathcal{A}$; these functions assign a symbol $x_{\boldsymbol{k}}$ to each point $\boldsymbol{k} \in \mathbb{Z}^d$. As we can see, this defines a family of topological dynamical systems, where the $\mathbb{Z}^d$-group action is defined by translations:

Definition 3.1 *The $\mathbb{Z}^d$-**full-shift** $\mathcal{A}^{\mathbb{Z}^d}$ is the set of all the aforementioned configurations with the prodiscrete topology, which can be described via the **shift metric**:*

$$d(x, y) := \sup(\{2^{-r} \;:\; x|_{[-r,r]^d} \neq y|_{[-r,r]^d}, r \in \mathbb{N}\} \cup \{0\}).$$

We define the following $\mathbb{Z}^d$-group action $\mathbb{Z}^d \overset{\sigma}{\curvearrowright} \mathcal{A}^{\mathbb{Z}^d}$ on the full-shift:

$$(\sigma_{\boldsymbol{m}}(x))_{\boldsymbol{k}} = x_{\boldsymbol{m}+\boldsymbol{k}}.$$

[1] In general, the word "configuration" refers to any function $K\colon U \to \mathcal{A}$, where $U \subseteq \mathbb{Z}^d$; this set U is called the **support** of the configuration. We shall place a special emphasis on the case where $|U| < \infty$, which will be described below.

This is called the ***shift action***[2]*, and the functions* $\sigma_{\boldsymbol{n}}$*, with* $\boldsymbol{n} \in \mathbb{Z}^d$ *are called* ***shift maps****. Any* $X \subseteq \mathcal{A}^{\mathbb{Z}^d}$ *that is both topologically closed and* ***σ-invariant*** *(that is,* $\sigma[X] = X$*), is called a* ***subshift****,* ***shift space*** *or simply* ***shift****.*

Subshifts are topological dynamical systems on their own, with the induced topology and the restriction of the group action to them. Interest in them arose originally from the fact that the associated group action is quite simple to describe, and yet they are sufficiently versatile to describe the behavior of seemingly much more complicated dynamical systems (e.g. billiards, geodesic flows, the baker's map, decimal expansions of numbers, etc.), with the difficulty in the description of the group action being transferred to appropriately describing the phase space, something that might be more tractable by combinatorial means. Thus, it is important to have a good grasp on the inner works of the topology of the full-shift and any subshift.

Definition 3.2 *A* ***pattern*** P *is a configuration of finite support, that is, given a finite set*[3] $U \Subset \mathbb{Z}^d$*, a pattern with support* $U =: \operatorname{supp}(P)$ *is any map* $P\colon U \to \mathcal{A}$*. We say that a pattern* Q *is a* ***translate***[4] *of* P *if* $\operatorname{supp}(Q) = \operatorname{supp}(P) + \boldsymbol{n}$ *and* $Q_{\boldsymbol{n}+\boldsymbol{k}} = P_{\boldsymbol{k}}$*.*

Given a pattern $P\colon U \Subset \mathbb{Z}^d \to \mathcal{A}$ *and some configuration* $K\colon V \subseteq \mathbb{Z}^d \to \mathcal{A}$ *(e.g. another pattern, or a point of* $\mathcal{A}^{\mathbb{Z}^d}$*), we say that* P *is* ***contained*** *in* K*, and write* $P \sqsubseteq K$*, if there is some translate* Q *of* P *such that* $K|_{\operatorname{supp}(Q)} = Q$*. The* ***language*** *of a shift space* X *is the set of all patterns that are contained in some* $x \in X$*, that is:*

$$\mathcal{L}(X) := \{x|_U \ : \ x \in X, U \Subset \mathbb{Z}^d\}.$$

We also write $\mathcal{L}_U(X)$ *for the set of all patterns of* $\mathcal{L}(X)$ *with support* U*. Similarly, given some* $x \in \mathcal{A}^{\mathbb{Z}^d}$*, we write* $\mathcal{L}(x)$ *(respectively,* $\mathcal{L}_U(x)$*) for the set of all patterns* P *(respectively, those with support* U*) with* $P \sqsubset x$*. Note that, if* $x \in X$*, then* $\mathcal{L}(x) \subseteq \mathcal{L}(X)$*.*

In the one-dimensional case, we often only use patterns $w = w_1 w_2 \dots w_n$ with support $\{1, 2, \dots, n\}$, which we call **words** (of length n), and we completely identify words by translation; we can develop the full theory of one-dimensional shift spaces with little to no change with this restriction. In this situation, we write $\mathcal{A}^*$ for the set of all words (including the **empty word** ε of length 0); this set is a monoid under concatenation, meaning that the following rule defines an associative operation for words:

$$(u \cdot w)_j = \begin{cases} u_j & \text{if } 1 \leq j \leq |u|, \\ w_{j-|u|} & \text{if } |u| < j \leq |u| + |w|, \end{cases}$$

where the notation $|w|$ refers to the length of the word w. Note that $\varepsilon \cdot w = w = w \cdot \varepsilon$.

In this one-dimensional case, the language $\mathcal{L}(X)$ of the shift space is **extensible**, meaning that every $w \in \mathcal{L}(X)$ is contained in a strictly larger word $w' \in \mathcal{L}(X)$, and **factorial**, i.e.

[2] Actually, the shift action may be defined for an arbitrary group G on $\mathcal{A}^G$. However, one must take into account that the group G is usually nonabelian, and thus one may not obtain a left group action directly; the most common definition used that compensates for this is given by $(\sigma_g(x))_h = x_{g^{-1}h}$, but others such as $(s_g(x))_h = x_{hg}$ are also common.

[3] We use the symbol $\Subset$ for "finite subset of".

[4] For most intents and purposes, we identify two patterns if one is a translate of the other.

every subword $w'' \sqsubseteq w \in \mathcal{L}(X)$ is also in $\mathcal{L}(X)$. The converse also holds: any collection of words with these two properties is the language of a subshift. Higher-dimensional analogues of these two notions exist, and characterize the set of patterns that appear in a shift space.

As stated before, the topology of shift spaces is (the one induced by) the product topology, taking $\mathcal{A}$ as a discrete space. This means that a sequence[5] $x^{(n)}$ of points of $\mathcal{A}^{\mathbb{Z}^d}$ converges to some $x^* \in \mathcal{A}^{\mathbb{Z}^d}$ if, for every finite $F \Subset \mathbb{Z}^d$, there is some $N(F)$ such that, for all $n > N(F)$ we have $x^{(n)}|_F = x^*|_F$. Thus, open sets for this topology must reflect this phenomenon: an open neighborhood for x^* in the shift space $X \subseteq \mathcal{A}^{\mathbb{Z}^d}$ would be the set of all $y \in X$ such that $x^*|_F = y|_F$, for some finite F. Formally:

Definition 3.3 *Given some pattern $P \colon U \Subset \mathbb{Z}^d \to \mathcal{A}$, the corresponding* ***cylinder*** *in a subshift X is a set of the form:*

$$[P] := \{x \in X \;:\; x|_U = P\}.$$

Cylinders form a base of the topology of a shift space. It is not hard to see that an open ball of radius 2^{-r} around some $x \in X$ is equal to the cylinder $[x|_{[-r,r]^d}]$; similarly, any cylinder can be seen as a finite union of open balls. We note, as well, that if a pattern Q is a translate of another pattern P, then $[Q]$ is the image of $[P]$ under some shift map $\sigma_{\boldsymbol{n}}$. This allows us to determine which subsets of $\mathcal{A}^{\mathbb{Z}^d}$ are subshifts in terms of cylinders, and, consequently, of patterns.

Indeed $P \not\sqsubseteq x$ if and only if $x \in \bigcap_{\boldsymbol{m} \in \mathbb{Z}^d} \sigma_{\boldsymbol{m}}[[P]]$, which is a closed, σ-invariant set and thus a subshift. It is not hard to see that, since closedness and σ-invariance are properties that transfer to intersections, any intersection of subshifts is a shift space as well. Thus, for any collection $\mathcal{F}$ of **forbidden patterns**, the set:

$$\mathsf{X}_{\mathcal{F}} := \{x \in \mathcal{A}^{\mathbb{Z}^d} \;:\; (\forall P \in \mathcal{F}) \colon P \not\sqsubseteq x\},$$

is a subshift; conversely, every subshift is of this form for some $\mathcal{F}$ (which may not be uniquely defined). In particular, we always have that $X = \mathsf{X}_{\mathcal{L}(X)^c}$.

Definition 3.4 ***Shifts of finite type*** *(SFT) are those of the form $X = \mathsf{X}_{\mathcal{F}}$ for which the set $\mathcal{F}$ of forbidden patterns may be chosen finite. If we can choose $\mathcal{F}$ so that every forbidden pattern has support $\{\mathbf{0}, \boldsymbol{e}_i\}$ for some element $\boldsymbol{e}_i$ of the canonical basis, we say that $\mathsf{X}_{\mathcal{F}}$ is a* ***nearest neighbor subshift*** *(or nnSFT).*

Proposition 3.5 *A subshift X is of finite type if, and only if, there exists some finite set $U \Subset \mathbb{Z}^d$ and a set of patterns $\mathcal{P}$ with support U such that, for every $x \in X$ and each $\boldsymbol{m} \in \mathbb{Z}^d$ we have that $x|_{\boldsymbol{m}+U}$ is a translate of some pattern in $\mathcal{P}$.*

For instance, one may take $\mathcal{P} = \mathcal{L}_U(X)$ for some sufficiently large finite set U. Note that in this case $X = \mathsf{X}_{\mathcal{L}_U(X)^c}$.

We shall be interested in subshifts where the shift action is **faithful** (i.e. not all points share a common period), as we defined in the chapter about groups and group actions. Periodic points

[5] We do not need the full generality of nets, since shifts are metric spaces.

are often important in our study, and thus we make some sub-classifications: a point $x \in \mathcal{A}^{\mathbb{Z}^d}$ is **strongly periodic** if $[\mathbb{Z}^d : \mathrm{Stab}(x)] < \infty$, and **weakly periodic** if $\mathrm{Stab}(x) \neq \{\mathbf{0}\}$. Note that strongly periodic points are also weakly periodic. A shift space is **weakly aperiodic** if it does not contain strongly periodic points, and **strongly aperiodic** if it does not contain weakly periodic points.

Since the topology of a shift space can be described wholly in terms of cylinders, and those in turn are entirely described by patterns, then it makes sense to describe certain topological and dynamical properties in terms of patterns and languages. First, remember that a topological dynamical system X is **transitive** if it has a point x^* whose orbit is dense in X, and **minimal** if every point has this property. In our context, transitivity means that the orbit of this point intersects every cylinder $[P]$, and thus $P \sqsubset x$. Indeed, we have that:

Proposition 3.6 *A subshift X is transitive if, and only if, there exists some $x^* \in X$ with $\mathcal{L}(x) = \mathcal{L}(X)$. Equivalently, for every $P \in \mathcal{L}(X)$, $P \sqsubset x$.*

Minimality is similarly described by the property that $\mathcal{L}(X) = \mathcal{L}(x)$ for every $x \in X$. However, we may characterize this differently:

Definition 3.7 *Let X be a $\mathbb{Z}^d$-subshift. We say that X is **recurrent** if, for every pattern $P \in \mathcal{L}(X)$, there exists some finite set $F(P) \Subset \mathbb{Z}^d$ such that, for every $x \in X$, $P \sqsubseteq x|_F(P)$.*

In practice, this means that there is a radius $r(P)$ (that depends on P) that ensures that we may find the pattern P at distance at most $r(P)$ from the origin. Since this applies to shifts of x as well, this also implies that, for all $\boldsymbol{n} \in \mathbb{Z}^d$, $P \sqsubseteq x|_{F(P)+\boldsymbol{n}}$ as well. Note that this obviously implies that X is minimal; the converse comes from the fact that, if this property were to be false, for a given pattern $P \in \mathcal{L}(X)$ one may find a sequence of points $x^{(n)} \in X$ such that no instance of P appears in $x^{(n)}$ at distance less than n from the origin. By compactness, we may find a point that does not contain P as the limit of some subsequence, proving that $\mathcal{L}(x) \neq \mathcal{L}(X)$. Thus, we have:

Theorem 3.8 *A $\mathbb{Z}^d$-subshift X is minimal if, and only if, it is recurrent.*

Note that, a priori, $r(P)$ does not have an obvious dependence on P beyond roughly increasing with the diameter of the support of this pattern. When $r(P)$ has an upper bound of the form $C \cdot \mathrm{diam}(\mathrm{supp}(P))$ for any pattern P, we say that X is a **linearly recurrent** subshift.

3.2. Sliding block codes

This work focuses on certain mappings that preserve some structure of a shift space. Our main point of interest are the following:

Definition 3.9 *Let $A \in \mathrm{GL}_d(\mathbb{Z})$, and X, Y be two $\mathbb{Z}^d$-subshifts (or, more generally, two $\mathbb{Z}^d$-topological dynamical systems). A continuous function $\varphi\colon X \to Y$ is said to be A-**equivariant** if:*

$$(\forall \boldsymbol{m} \in \mathbb{Z}^d)\colon \varphi \circ \sigma_{\boldsymbol{m}} = \sigma_{A\boldsymbol{m}} \circ \varphi,$$

and, in particular, when $A = I_d$, we call φ a ***sliding block code***[6] *(or* ***block map****). When φ is an homeomorphism and $A = I_d$, we say that it is a* ***conjugacy*** *(or sometimes* ***isomorphism****) and that X and Y are* ***conjugate subshifts****, writing $X \cong Y$.*

In the particular case where $X = Y$, we refer to φ as either an ***automorphism***[7]*, if $A = I_d$, or an* ***extended symmetry*** *in the general case (thus, automorphisms are extended symmetries, but the converse is usually not true).*

It can be proved that the inverse function of a bijective sliding block code is also a sliding block code, which justifies the use of the word isomorphism in this situation; however, this word is more often used in a measurable context. This also implies that the collection of all automorphisms is a group, which we shall denote $\mathrm{Aut}(X, \mathbb{Z}^d)$ or $\mathrm{Aut}(X, \sigma)$ (some authors use $\mathcal{S}(X)$ instead). Extended symmetries form a group as well, which we shall denote as $\mathrm{Sym}(X, \mathbb{Z}^d)$ or $\mathrm{Sym}(X, \sigma)$ (or[8] $\mathcal{R}(X)$). General A-equivariant maps are described in further detail by de Neymet [32].

Extended symmetries (and, more generally, any A-equivariant map) are easily characterized by the following result, which we cite in full generality [13, 24, 71] and explains the name "sliding block code":

Theorem 3.10 (Generalized Curtis–Hedlund–Lyndon theorem, or CHL) *Let X, Y be two subshifts over alphabets $\mathcal{A}_X, \mathcal{A}_Y$. A continuous map $\varphi\colon X \to Y$ is A-equivariant if, and only if, there exists some finite set $U \Subset \mathbb{Z}^d$ (called* ***memory set****) and a function $\Phi\colon \mathcal{A}_X^U \to \mathcal{A}_Y$ (****local function***[9]*) such that:*

$$(\forall \boldsymbol{m} \in \mathbb{Z}^d)\colon (\varphi(x))_{A\boldsymbol{m}} = \Phi(x|_{\boldsymbol{m}+U}).$$

Note that here, if P is a pattern with support U and Q is any translate of P, we define $\Phi(Q) := \Phi(P)$.

When $A = I_d$ we recover the classic **Curtis–Hedlund–Lyndon theorem**, which is one of the central results from symbolic dynamics. A proof for the one-dimensional case when $A = I_1 = 1$ may be found in the book by Lind and Marcus [71]; very little work needs to be done to convert this proof into one that works for the completely general result (see the paper by Baake, Roberts and Yassawi [13] and the book on cellular automata by Ceccherini-Silberstein and Coornaert [24]). We note that we may always assume the memory set U to be of the form $[-r, r]^d$ for some r; the least possible r is called the *radius* of φ. When φ is an automorphism of radius 0, we refer to it as a *letter swap* or *relabeling map*. In the one-dimensional case, one may be more specific, and note that the memory set U is contained in an interval of the form $[-m, n]$ with $m, n \geq 0$; the values m and n are called **memory** and **anticipation**, respectively. Note that at least one of them must equal the radius r.

[6] This term is reserved for the situation where both X and Y are shift spaces, due to the Curtis–Hedlund–Lyndon theorem introduced below.

[7] Some authors follow the Smale convention in which every homeomorphism of X to itself is called an automorphism, irregardless of whether it commutes with the shift action or not; these authors use the word **symmetry** for what we are designating as "automorphisms".

[8] This is a remnant of the one-dimensional case where extended symmetries that are not automorphisms necessarily satisfy the equation $\varphi \circ \sigma = \sigma^{-1} \circ \varphi$ and are thus called **reversors**.

[9] We use the convention from [71], that uses an uppercase version of the symbol used to denote an A-equivariant map to refer to its local function.

Surjective sliding block codes $\varphi\colon X \to Y$ (or, more generally, surjective equivariant maps between two topological dynamical systems) are called **factor maps**, and in this situation we say that Y is a **factor** of X. While a subshift conjugate to a shift of finite type is also of finite type, this is not true for factors; thus, when $\varphi\colon X \to Y$ is a factor map and X is a SFT, we say that Y is a **sofic shift**.

Since sliding block codes preserve the shift action and are continuous, many topological and dynamical properties are preserved under them. For example, it is very easy to check that:

Proposition 3.11 *The image of a $\boldsymbol{p}$-periodic point under an A-equivariant map φ is an $(A\boldsymbol{p})$-periodic point. Thus,* $\operatorname{Stab}(\varphi(x)) \supseteq A \cdot \operatorname{Stab}(x)$.

Note that, when φ is not injective, it may be the case that $\varphi(x)$ has additional periods. For instance, the factor map $\varphi\colon \{a,b,c,d\}^{\mathbb{Z}} \to \{0,1\}^{\mathbb{Z}}$ given by $a, c \mapsto 0$, $b, d \mapsto 1$ maps the 4-periodic point[10] $(abcd)^\infty$ to the 2-periodic point $(01)^\infty$, and thus there is a strict inclusion $\operatorname{Stab}((abcd)^\infty) = 4\mathbb{Z} \subset 2\mathbb{Z} = \operatorname{Stab}((01)^\infty)$.

Proposition 3.12 *If $\varphi\colon X \to Y$ is a factor map (or, more generally, a surjective A-equivariant map) and $x \in X$ is a transitive point, then so is $\varphi(x) \in Y$ (i.e. the image of a topologically transitive subshift under a factor map is topologically transitive).*

Other properties that depend only on convergence properties and the shift action (such as asymptotic pairs) are preserved as well, maybe after taking into account the matrix A. The following theorem is sometimes useful in the characterization of sliding block codes:

Theorem 3.13 *Let $\varphi\colon X \to Y$ any sliding block code between two shift spaces. Then, there is another subshift $\tilde{X}$ which is conjugate to X via the map ψ and a radius* 0 *sliding block code $\tilde{\varphi}\colon \tilde{X} \to Y$ such that $\varphi = \tilde{\varphi} \circ \psi$, as in the below diagram:*

$$\begin{tikzcd}
X \arrow[r, "\psi"] \arrow[d, "\varphi"'] & \tilde{X} \arrow[dl, "\tilde{\varphi}"] \\
Y &
\end{tikzcd}$$

Something similar holds for general A-equivariant maps between subshifts. The key here is that, if U is a memory set for φ, one may create $\tilde{X}$ as a new subshift whose alphabet is (in bijection with) $\mathcal{L}_U(X)$, and defining forbidden patterns that only allow two symbols of this new alphabet to be adjacent if their overlap is consistent (i.e. if $(\boldsymbol{m} + U) \cap (\boldsymbol{n} + U) \neq \varnothing$, the two patterns $P, Q \in \mathcal{L}_U(X)$ can appear in positions $\boldsymbol{m}$ and $\boldsymbol{n}$ in some point of $\tilde{X}$ if the symbols of P and Q in this intersection are the same), and no forbidden pattern of the original shift space appears. This construction is often called the **higher block shift**.

It is important to note that compositions and inverses of automorphisms are also automorphisms, and the same holds for extended symmetries. Thus, they define special subgroups

[10] The notation w^∞ here refers to the point of $\mathcal{A}^{\mathbb{Z}}$ obtained by concatenating infinitely many copies of the word w. More complicated notations such as $w_1^\infty.w_2w_3^\infty$ are also in use, where the dot indicates the position of the zero coordinate.

of the set Homeo(X) of all self-homeomorphisms of the topological space X, called respectively the **automorphism group** $\mathrm{Aut}(X,\mathbb{Z}^d)$ and **extended symmetry group** $\mathrm{Sym}(X,\mathbb{Z}^d)$. In group-theoretic terms, the automorphism group $\mathrm{Aut}(X,\mathbb{Z}^d)$ equals the **centralizer** of the subgroup $\langle\sigma\rangle = \{\sigma_{\boldsymbol{n}} : \boldsymbol{n} \in \mathbb{Z}^d\}$ of all shift mappings in Homeo(X), while the group of extended symmetries $\mathrm{Sym}(X,\mathbb{Z}^d)$ is the **normalizer** of $\langle\sigma\rangle$ in Homeo(X). Hence, passing from $\mathrm{Aut}(X,\mathbb{Z}^d)$ to $\mathrm{Sym}(X,\mathbb{Z}^d)$ is not an arbitrary generalization, as the latter has a natural algebraic definition in itself. Furthermore, both groups are **conjugacy invariants**, that is, when two subshifts are conjugate, their corresponding automorphism and extended symmetry groups are isomorphic, and thus "the same" from a group-theoretical viewpoint.

3.3. Substitutions

Most of the subshifts we shall deal with in what follows are **substitutive subshifts**. They are described by a **rectangular substitution** $\theta\colon \mathcal{A} \to \mathcal{A}^R$, where $R = [\mathbf{0}, \boldsymbol{s}-\mathbf{1}] = \prod_{i=1}^{d}[0, s_i-1]$, which is a map that associates to every symbol in $\mathcal{A}$ a rectangular pattern of symbols in the same alphabet. We extend θ to arbitrary patterns or configurations by concatenation in every direction (e.g. if P is a pattern with a and b are two symbols adjacent along the direction $\boldsymbol{e}_1$, $\theta(P)$ will contain the patterns $\theta(a)$ and $\theta(b)$, adjacent along the same direction) and then we define:

Definition 3.14 *Given a rectangular substitution $\theta\colon \mathcal{A} \to \mathcal{A}^R$, the associated* ***substitutive subshift*** *is given by:*

$$\mathsf{X}_\theta{:=}\{x \in \mathcal{A}^{\mathbb{Z}^d} : \textit{for any finite } U \Subset \mathbb{Z}^d,\ \textit{there are some } k \in \mathbb{Z}, a \in \mathcal{A} \textit{ such that } x|_U \sqsubseteq \theta^k(a)\}.$$

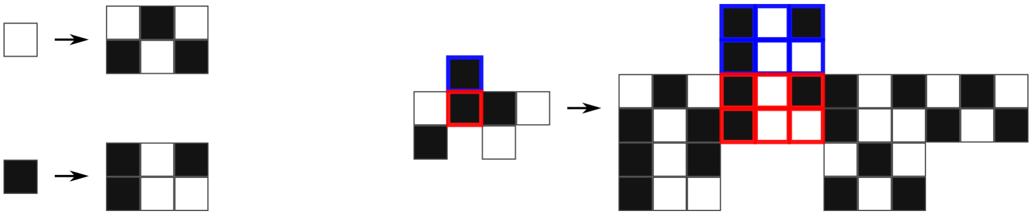


Figure 3.1: A rectangular substitution extends to arbitrary patterns by concatenation, i.e. preserving adjacencies.

In the one-dimensional case, we do not have the restriction of all words $\theta(a), a \in \mathcal{A}$ to have the same length, and thus it makes sense to think of a substitution as a function $\theta\colon \mathcal{A} \to \mathcal{A}^*\setminus\{\varepsilon\}$. For instance, the **golden mean substitution** (or **Fibonacci substitution**) is given by the map:

$$\begin{aligned}\theta\colon 0 &\mapsto 01\\ 1 &\mapsto 0,\end{aligned}$$

so that the sequence of patterns obtained by iterated substitution are of the form:

$$0 \mapsto 01 \mapsto 010 \mapsto 01001 \mapsto 01001010 \mapsto 0100101001001 \mapsto \ldots$$

that is, $w_0 = 0, w_1 = 01$, and $w_{n+2} = w_{n+1}w_n$ for $n \geq 0$, and a point $x \in \{0,1\}^*$ belongs to the associated subshift if there are arbitrarily large central patterns that appear in this sequence of words. For $d = 1$, when the restriction of lengths is respected, we talk of a **constant length substitution** (of length ℓ) and see rectangular substitutions as a generalization of this particular class of one-dimensional substitutions. We recommend Kůrka's book [64] as an introduction to the one-dimensional theory of substitutive subshifts, and the work by Frank [41] as an entrance point for the multidimensional case.

Since we extended the function θ to arbitrary patterns or configurations, expressions such as $\theta^k(a)$ are well-defined by induction (e.g. θ^2 may be seen as a map $\mathcal{A} \to \mathcal{A}^{R^{(2)}}$ where $R^{(2)} = \prod_{i=1}^{d}[0, s_i^2 - 1]$, obtained by applying θ to every pattern $\theta(a)$ and concatenating appropriately) and define substitutions as well. This allows us to define the following important property:

Definition 3.15 *A substitution θ is* ***primitive*** *if, for some $k \geq 1$, every pattern $\theta^k(a)$ contains every symbol in the alphabet.*

In this case, it is easy to verify that $\mathsf{X}_\theta = \mathsf{X}_{\theta^k}$ for any k, and thus we can replace θ by a suitable power whenever appropriate. Using this characterization, we can also prove the following:

Theorem 3.16 *If θ is primitive, X_θ is a minimal subshift.*

Due to these properties, in what follows we shall assume that every substitution we meet is primitive, as the theory of non-primitive substitutions and the techniques used to study the corresponding subshifts is quite different [75]. The name "primitive" comes from the corresponding characterization via matrices:

Definition 3.17 *Let $\theta\colon \mathcal{A} \to \mathcal{A}^R$ be a rectangular substitution, where $\mathcal{A} = \{a_1, \dots, a_m\}$. The corresponding* ***substitution matrix*** *is a $m \times m$ matrix A with non-negative integer entries, where $A_{ij} = k$ when the pattern $\theta(a_i)$ has exactly k instances of the symbol a_j.*

It is easy to see that, if A is the substitution matrix of the substitution θ, then A^n is the corresponding matrix of the iterated substitution θ^n. We can use, then, this matrix to characterize the properties of the substitution.

Definition 3.18 *A $m \times m$ matrix $A \in \mathbb{M}_m(\mathbb{R})$ with nonnegative entries is* ***irreducible*** *if, for every $1 \leq i, j \leq m$, there is some power k such that $A^k_{i,j} > 0$. If there is some $k > 0$ such that all entries of A^k are strictly positive, then we say that A is* ***primitive****.*

Proposition 3.19 *A substitution θ is primitive if, and only if, its associated substitution matrix is primitive.*

Since we often deal with iterated substitutions and automorphisms that a priori might have positive radius, the following two notations prove useful:

$$R^{(r)} := [\mathbf{0}, \boldsymbol{s}^r - \mathbf{1}] = \prod_{i=1}^{d}[0, s_i^r - 1], \qquad U^{\circ r} := \{\boldsymbol{k} \in U \ : \ \boldsymbol{k} + [-r, r]^d \subseteq U\}.$$

Indeed, if θ is a rectangular substitution of shape R, then the shape of θ^k is $R^{(k)}$. Similarly, if $f\colon X \to Y$ is a sliding block code with radius R and $x, y \in X$ satisfy the equality $x|_U = y|_U$, then $f(x)|_{U^{\circ r}} = f(y)|_{U^{\circ r}}$. Something similar holds for general A-equivariant maps.

As a substitution θ can be extended to any configuration, it naturally induces a mapping $\theta_\infty\colon \mathcal{A}^{\mathbb{Z}^d} \to \mathcal{A}^{\mathbb{Z}^d}$. It can be verified [41, 64] that there exists a θ_∞-periodic point $x^* \in \mathcal{A}^{\mathbb{Z}^d}$, i.e. $\theta_\infty^k(x^*) = x^*$, such that $\mathsf{X}_\theta = \overline{\mathrm{Orb}(x^*)}$; any such point is entirely determined by its central pattern $x^*|_{\{-1,0\}^d}$, which must be some element of $\mathcal{L}_{\{0,1\}^d}(\mathsf{X}_\theta)$; we call this pattern the **seed** of x^*. This implies that, for any $k \in \mathbb{N}$, we may see x^*, and thus any point $x \in \mathsf{X}_\theta$ (due to minimality), as a concatenation of patterns of the form $\theta^k(a), a \in \mathcal{A}$; it turns out that, in a specific sense, this can be done in only one possible way:

Theorem 3.20 (Recognizability) *Let $\theta\colon \mathcal{A} \to \mathcal{A}^R$ be a primitive rectangular substitution and suppose the associated substitutive subshift X_θ has a faithful shift action. For each $x \in \mathsf{X}_\theta$, and every $k > 0$, there exist unique $\boldsymbol{m}_k \in R^{(k)}$ and $y_k \in \mathsf{X}_\theta$ such that $x = \sigma_{\boldsymbol{m}_k}(\theta_\infty^k(y_k))$. Furthermore, for any $\ell > k$, we have $\boldsymbol{m}_k \equiv \boldsymbol{m}_\ell$ (mód $\boldsymbol{s}^k$), with this congruence being taken componentwise.*

This is a special case of a **recognizability property**, which implies that we can "undo" the substitution to a certain extent, that is, that we can recover a pattern P from $\theta(P)$ except maybe at its border. See the works by Mossé [79] and Solomyak [94] for reference; additional details can be found in the work by Frank [41], and the books by Kůrka [64], Pytheas Fogg [38] and Baake and Grimm [6]. When dealing with infinite configurations, this recognizability property translates into what is known as a **box structure** (see Olli [83]), which is a series of "grids" (sublattices of $\mathbb{Z}^d$, as defined further below) of increasing size, which mark the positions of the patterns $\theta^k(a), a \in \mathcal{A}$ in which a given point $x \in \mathsf{X}_\theta$ decomposes, for every value of k. This can be seen in Figure 3.2.

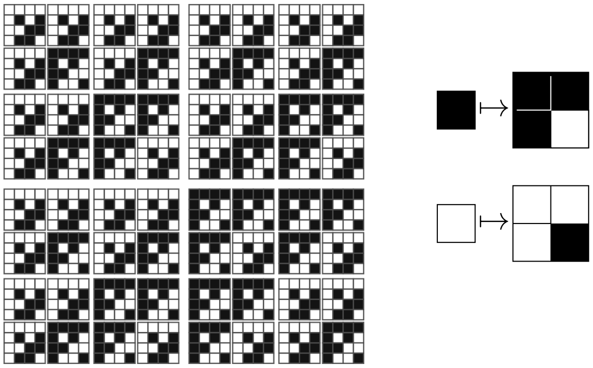


Figure 3.2: $2^n \times 2^n$ grids associated with the iterates of a primitive substitution θ in a point from a substitutive subshift. The corresponding substitution is indicated on the right.

The recognizability theorem is often stated as the existence of a factor map from X_θ to a product of odometers (a special case of a generalized odometer, as defined at the end of Chapter 1) $\mathbb{Z}_{\boldsymbol{s}} = \mathbb{Z}_{s_1} \times \cdots \times \mathbb{Z}_{s_k}$ [41], which maps a point $x \in \mathsf{X}_\theta$ to the corresponding sequence of shifts $[\boldsymbol{m}_1, \boldsymbol{m}_2, \dots] \in \mathbb{Z}_{\boldsymbol{s}}$. Under specific circumstances, this product of odometers is the **maximal equicontinuous factor** (MEF) of X_θ and, at least in the one-dimensional case, $\mathbb{Z}_{\boldsymbol{s}}$ is always a finite index subgroup of this MEF. However, we shall not use this version of the result except maybe for some passing remarks below. Regardless, we shall mention that the main obstacle for this product of odometers to actually be the MEF of the subshift lies in the following:

Definition 3.21 *Let $\theta\colon \mathcal{A} \to \mathcal{A}^\ell$ be a one-dimensional constant length substitution. We say that the* ***height*** *of θ is the following value:*

$$h = \text{máx}\{n \geq 1 \,:\, \gcd(n, \ell) = 1 \text{ and } n \mid \gcd(\{m \in \mathbb{N} \,:\, u_m = u_0\})\},$$

where u is a periodic point of θ. This value is independent of the chosen fixed point u.

If the substitution θ has height $h > 1$, we can partition the alphabet $\mathcal{A}$ into h disjoint subsets $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_h$ in such a way that every symbol from $\mathcal{A}_{j+1}$ appears right next to a symbol from $\mathcal{A}_j$ (where the indexing values j are taken modulo h). It can be shown that the MEF of the subshift X_θ is the odometer $\mathbb{Z}_\ell$ whenever the height h equals 1.

We now introduce a classification for certain kinds of substitutions due to their specific properties:

Definition 3.22 *Given a rectangular substitution $\theta\colon \mathcal{A} \to \mathcal{A}^R$ and a $\boldsymbol{k} \in R$, the $\boldsymbol{k}$-th* ***column*** *of θ is the mapping $\theta_k\colon \mathcal{A} \to \mathcal{A}$, $a \mapsto \theta(a)_{\boldsymbol{k}}$. We say that θ is* ***bijective*** *if all of its columns are bijections; otherwise, if for some power m we have that $\theta^m(a)_{\boldsymbol{k}} = \theta^m(b)_{\boldsymbol{k}}$ for all $a, b \in \mathcal{A}$ and some $\boldsymbol{k} \in R^{(m)}$, we say that θ has a* ***coincidence****.*

Remark When $|\mathcal{A}| = 2$, a rectangular substitution is either bijective or has a coincidence. For alphabets with larger cardinality, intermediate cases can be found, where θ does not satisfy either definition.

Since $\mathsf{X}_\theta = \mathsf{X}_{\theta^k}$ whenever θ is primitive, we note that it is often useful to replace θ with a suitable power. In particular, when θ is bijective, we may choose a k that is large enough for a given column $\theta_{\boldsymbol{k}}$ to be the identity. Thus, we may always assume, for instance, that $\theta(a)$ has the symbol a at every corner of the d-dimensional rectangle; this implies that every point x that is periodic under θ_∞ is a θ_∞-fixed point, which makes easier to compute certain properties of the associated shift space.

3.4. Entropy

Entropy is a very important invariant in several branches of the theory of dynamical systems, both in its ergodic and topological versions. We shall give some brief notes about topological entropy in what follows.

Definition 3.23 *Given a $\mathbb{Z}^d$-subshift X, the* ***topological entropy*** *of X is the following value:*

$$h_{\text{top}}(X) := \lim_{n\to\infty} \frac{\log(|\mathcal{L}_{[1,n]^d}(X)|)}{n^d} = \inf_{n\geq 1} \frac{\log(|\mathcal{L}_{[1,n]^d}(X)|)}{n^d}.$$

This definition can be generalized to general amenable groups via the use of so-called Følner nets, where it can be proven that the equivalent limit converges and equals the infimum; see

Ceccherini-Silberstein and Coornaert [24] for details. We will not be needing the full generality of Følner nets in what follows. Note that we do not specify the base of the logarithm, as it is irrelevant to our purposes; for simplicity, we will assume the logarithm to be natural, but in other branches of mathematics (e.g. computer science) it makes sense to take logarithms in base 2 (or, more generally, base $|\mathcal{A}|$).

Entropy is a quantity that depends only on the conjugacy class of a given subshift, i.e. $X \cong Y$ implies that $h_{\text{top}}(X) = h_{\text{top}}(Y)$; thus, computing entropy gives us a great deal of information on the problem of deciding whether two shifts are "the same" from a topological dynamics viewpoint. Indeed, we have that:

Proposition 3.24 *If $f\colon X \to Y$ is a factor map (or, more generally, a surjective A-equivariant map), then $h_{\text{top}}(X) \geq h_{\text{top}}(Y)$.*

We have a similar property for inclusions:

Proposition 3.25 *If $X \subseteq Y$ are both $\mathbb{Z}^d$-subshifts, then $h_{\text{top}}(X) \leq h_{\text{top}}(Y)$.*

Note that we could have equality even when there is a strict inclusion. In the one-dimensional case, there are some conditions that ensure that a strict inclusion implies a strict inequality, but these conditions fail for $d > 1$, which is our main case of interest, so we will not go into detail.

Sometimes it is good to go into some finer detail. The following definition, while not a conjugacy invariant, allows us to study the behavior of systems for which entropy is too rough a measurement:

Definition 3.26 *Let X be a $\mathbb{Z}^d$-subshift. The* ***complexity function*** *of X is the function $p_X\colon \mathbb{N}^d \to \mathbb{N}$ given by:*

$$p_X(\boldsymbol{n}) := |\mathcal{L}_{[\boldsymbol{1},\boldsymbol{n}]}(X)|,$$

where $[\boldsymbol{1},\boldsymbol{n}] = \prod_{k=1}^{d}[1,n_k]$, $\boldsymbol{n} = (n_1,\ldots,n_d)$.

Thus, if $h_{\text{top}}(X) = h$, then $p_X(n,\ldots,n) \approx ce^{hn^d}$ for large values of n. The real importance of this function comes to be noticeable when X is a shift of entropy 0, as $p_X(n_1,\ldots,n_d)$ may have wildly different behavior for different subshifts even if they have the same entropy.

Definition 3.27 *A subshift X has* ***linear complexity***[11] *if there exists some constant $C > 0$ such that:*

$$p_X(n_1,\ldots,n_d) \leq Cn_1\cdots n_d.$$

Thus, sublinear compexity means that there is very low variety among the patterns that can be found in $\mathcal{L}(X)$, as the number of possibilities for some $P \in \mathcal{L}_U(X)$ is roughly proportional to the cardinality of U; to contrast, in a full shift over an alphabet $\mathcal{A}$, we would have $|A|$ possibilities for every coordinate in U, and thus $|\mathcal{L}_U(\mathcal{A}^{\mathbb{Z}^d})| = |A|^{|U|}$, an exponential growth. Note that this also serves as a proof that, for the full shift in n symbols, the entropy is $\log(n)$.

[11] Some authors use **sublinear complexity** for what we call "linear complexity" here. This is because of consistency with other branches of mathematics that deal with growth orders of functions.

Proposition 3.28 *Substitutive subshifts have linear complexity.*

Corollary 3.29 *Any substitutive subshift must have entropy $h_{\text{top}}(\mathsf{X}_\theta) = 0$.*

3.5. Sets of points and tilings

In this section, we briefly introduce some basic notions from tiling theory, as, while the setting is by nature different to the one of symbolic dynamics, it shares a big part of its metodology and approaches. Our main references for the following sections are the books by Baake and Grimm [6] and Sadun [89]. We start by discussing some basic definitions for sets of points:

Definition 3.30 *A set of points $\Lambda \subseteq \mathbb{R}^d$ is **discrete** if every point of Λ is isolated, i.e. every $x \in \Lambda$ has an open neighborhood $U_x \subseteq \mathbb{R}^d$ such that $\Lambda \cap U_x = \{x\}$. If there is some open neighborhood U of 0 such that $U_x = U + x$ for every $x \in \Lambda$, we say that this set is **uniformly discrete**.*

We may allow sets of points to be labelled, that is, we assign to each point in a point set Λ a symbol in some alphabet $\mathcal{A}$ (indeed, one may think of such a labelled point set as a function $\Gamma\colon \Lambda \to \mathcal{A}$ where Λ is a point set in the above sense, or as a collection of point sets $\Gamma_a = \Gamma^{-1}[\{a\}]$, one for each $a \in \mathcal{A}$). The notions defined below for point sets extend directly to labelled point sets.

Definition 3.31 *A set of points $\Lambda \subseteq \mathbb{R}^d$ is r**-dense** if every point of $\mathbb{R}^d$ is at distance at most r from some point of Λ, i.e. if $\Lambda + \overline{B(\mathbf{0}, r)} = \mathbb{R}^d$. We say that Λ is **relatively dense** if it is r-dense for some value of $r > 0$.*

Definition 3.32 *A **Delone set** is a set $\Lambda \subset \mathbb{R}^d$ that is both uniformly discrete and relatively dense. A Delone set is called a (r, R)**-set** if, for every $\boldsymbol{x} \neq \boldsymbol{y} \in \Lambda$ we have $r \leq d(\boldsymbol{x}, \boldsymbol{y}) \leq R$.*

We may assign a topology to the collection of all point sets $2^{\mathbb{R}^d}$ (not necessarily Delone) which comes from the generalization of the shift metric to this continuous setting. We say that two point sets Λ, Δ are **ε-close** if there are vectors $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \overline{B(\mathbf{0}, \varepsilon/2)}$ such that:

$$(\Lambda + \boldsymbol{v}_1) \cap \overline{B(\mathbf{0}, 1/\varepsilon)} = (\Delta + \boldsymbol{v}_2) \cap \overline{B(\mathbf{0}, \varepsilon)},$$

that is, the two patterns are equal in a large central patch, up to a small translation of size at most ε. The **tiling distance** between both point sets is the infimum of all ε such that Λ and Δ are ε-close, or 1 if no such ε exists or is greater than 1.

Note that this metric makes the translation functions $\alpha_{\boldsymbol{v}}\colon 2^{\mathbb{R}^d} \to 2^{\mathbb{R}^d}$ given by $\alpha_{\boldsymbol{v}}(\Lambda) = \Lambda + \boldsymbol{v}$ continuous, and thus this defines a group action $\mathbb{R}^d \overset{\alpha}{\curvearrowright} 2^{\mathbb{R}^d}$; those constitute a continuous analogue to the shift maps from symbolic dynamics. Indeed, we may define then a topological dynamical system analogous to a subshift, comprised of a closed collection of point sets under the above topology, which is also closed under translations by some subgroup (often not the whole group) of $\mathbb{R}^d$.

Patterns have an analogue for point sets, which consists of the set of points over a bounded region of the plane (analogue to the finite support of patterns in the symbolic dynamics context):

Definition 3.33 *A **cluster** of shape K (where K is a compact neighborhood of $\mathbf{0}$) on a point set Λ is any set of the form $\Lambda \cap (K + \boldsymbol{p})$, with $\boldsymbol{p} \in \Gamma$.*

As with patterns in the context of symbolic dynamics, we identify two clusters of shape K if one is a translate of the other.

Definition 3.34 *A discrete point set Λ has **finite local complexity** if, for every neighborhood K of $\mathbf{0}$, there is a finite number of clusters of shape K in Λ, up to translation.*

One comparatively very simple type of point set we shall encounter often is the following:

Definition 3.35 *A **lattice** in $\mathbb{R}^d$ is a subgroup $\Gamma \leq \mathbb{R}^d$ that is both discrete and **co-compact** (that is, the quotient $\mathbb{R}^d/\Gamma$ is compact). Equivalently, Γ is a lattice if there are d linearly independent vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d \in \mathbb{R}^d$ such that:*

$$\Gamma = \{n_1\boldsymbol{b}_1 + \cdots + n_d\boldsymbol{b}_d \ : \ n_1, \ldots, n_d \in \mathbb{Z}\}.$$

The second definition is the one that will be useful for our purposes; however, we state both as the first definition applies to more general groups. An example of lattice we have already encountered is the **Minkowski embedding** of a field K of index d over $\mathbb{Q}$ in $\mathbb{R}^d$; see Chapter 2. Evidently, a lattice is a Delone set.

In the context of point sets, it is important to determine “how often” points of Γ may be encountered. The following quantities allow us to measure the sparseness of a point set:

Definition 3.36 *The **upper natural density** of a point set Λ is the following quantity:*

$$\overline{\text{dens}}(\Lambda) := \text{lím sup}_{n\to\infty} \frac{|\Lambda \cap \overline{B}(\mathbf{0}, r)|}{\mu(\overline{B}(\mathbf{0}, r))},$$

*where μ is the standard Lebesgue measure in $\mathbb{R}^d$. The **lower natural density** $\underline{\text{dens}}(\Lambda)$ is defined analogously with the limit inferior. When both values coincide, we say that Λ has **natural density** equal to $\text{dens}(\Lambda) = \overline{\text{dens}}(\Lambda) = \underline{\text{dens}}(\Lambda)$.*

Note that natural density is defined via symmetric balls around the origin. This is important, as the above limits may change when using other kinds of averaging sets (e.g. more general Følner or van Hove nets). For instance, the set of square-free integer numbers $V^{(2)} \subset \mathbb{Z}$ can be seen to have natural density $6/\pi^2$, as, intuitively, the subset of integers from $[-r, r]$ that are not divisible by p^2 is approximately $2r(1 - p^{-2})$ for every prime p; then, the product of all $(1 - p^{-2})$ for $p \in \mathbb{P}$ converges to $1/\zeta(2) = 6/\pi^2$, from where the result may be obtained. However, it can be proved that $V^{(2)}$ has gaps of arbitrarily large size; thus, taking averaging sets of increasing diameter that happen to be contained in these gaps will result in a computed density of 0.

This observation is important because density in number-theoretical applications is often defined for the positive integers only, using $[1, r]$ instead of $[-r, r]$ as averaging sets, and hence a subset of $\mathbb{N}$ will have different densities for both definitions. We note, however, that the shape itself is not than important under some reasonable constraints; for instance, the following limite superior is also equal to the upper natural density of Λ:

$$\overline{\operatorname{dens}}(\Lambda) = \operatorname{lím\,sup}_{n\to\infty} \frac{|\Lambda \cap [-r,r]^d|}{(2r)^d},$$

A quick computation shows immediately that:

Definition 3.37 *The natural density of a lattice Γ given by the d linearly independent vectors $\boldsymbol{b}_1, \dots, \boldsymbol{b}_d$ is $1/V$, where V is the volume of the parallelogram whose edges are given by these vectors. If $\boldsymbol{b}_i = (b_{i,1}, \dots, b_{i,d})$, then V is the determinant of the matrix $[b_{i,j}]_{i,j}$, up to a sign.*

Given any point set Λ and a cluster C of Λ of shape K, we may define numbers that measure how often we may find translates of C in K. The following quantities accomplish this purpose:

Definition 3.38 *For a point set Λ and $C = \Lambda \cap (K + \boldsymbol{p}), \boldsymbol{p} \in \Lambda$ a cluster of shape K, define the set:*

$$L_\Lambda(C) := \{\boldsymbol{q} \in \Lambda \ : \ \Lambda \cap (K + \boldsymbol{q}) \text{ is a translate of } C\}.$$

The ***upper frequency*** *of C in Λ is the upper natural density of $L_\Lambda(C)$. The* ***lower frequency*** *is defined similarly, and we speak just of* ***frequency*** *if both quantities coincide.*

This definition extends naturally to the discrete setting of subshifts, where for a given $x \in \mathcal{A}^{\mathbb{Z}^d}$ we define the (upper, lower) frequency of a pattern P as the (upper, lower) natural density of the subset of $\mathbb{Z}^d$ of those points $\boldsymbol{p}$ where $x|_{\operatorname{supp}(P)+\boldsymbol{p}}$ is a translate of P. If, for a given subshift X, this is a proper frequency that does not depend on the chosen $x \in X$, then this may be used to define a measure on X which is preserved by shift maps, which is called the **frequency measure**; examples of this situation include, e.g. Sturmian subshifts. Since automorphisms need to preserve this measure, it is an useful tool in the study of which local functions effectively define automorphisms or extended symmetries of a shift space.

We are now in a position to define tiling spaces, which may be seen as a particular subcase of labelled point sets. In the most general sense, a **tiling** of $\mathbb{R}^d$ is a decomposition of $\mathbb{R}^d$ as a nonempty union of nonempty subsets (called **tiles**) of $\mathbb{R}^d$. However, we shall impose some limitations on the kinds of subsets of $\mathbb{R}^d$ that shall be tiles. The limitations often imposed are the following (which, while somewhat redundant, are made explicit for clarity):

- Each tile equals the closure of its interior (this, in particular, implies that tiles must have nonempty interior).
- The boundary of a tile has Lebesgue measure 0.
- Two tiles intersect only in their boundary, i.e. the interiors of two different tiles are disjoint.
- Tiles are bounded.

For our purposes, we can assume tiles to be polytopes (polygons, polyhedra, etc.), and that they meet face-to-face, that is, the intersection of two tiles equals a shared face (e.g. a common vertex, edge, etc.). Furthermore, we shall only be interested in the situation where there is a finite collection $\mathcal{T}$ of subsets of $\mathbb{R}^d$ satisfying the above restrictions, called **prototiles**, such that every tile is a translate of some prototile; note that, as in with point sets, we may allow "labelled" or "decorated" prototiles (in which we assign symbols of an alphabet to each prototile, which we may represent via arrows indicating orientation, colors, etc.), to allow different prototiles with the same shape.

Definition 3.39 *A* ***simple tiling*** *is a decomposition of $\mathbb{R}^d$ as a union of denumerably many polygonal sets (tiles) following the restrictions above, (they equal the closure of their interior, are bounded polytopes that meet face-to-face, etc.), and such that there exists some set $\mathcal{T}$ of (possibly labelled) finitely many subsets of $\mathbb{R}^d$, called prototiles, for which every tile is a translate of an element of $\mathcal{T}$.*

Thus, for every tiling T and every prototile $\tau \in \mathcal{T}$ the set:

$$\Lambda_\tau(T) := \{\boldsymbol{p} \in \mathbb{R}^d \ : \ \boldsymbol{p} + \tau \text{ is a tile of } T\},$$

is a discrete point set. Hence, the collection of all $\{\Lambda_\tau\}_{\tau\in\mathcal{T}}$ may be seen as a labelled Delone point set, which, together with the set of prototiles $\mathcal{T}$, entirely determines the tiling T, and thus the notions of tiling distance, cluster, finite local complexity, etc. translate immediately to the realm of tilings.

Definition 3.40 *A* ***tiling space*** *Ω is a set of tilings (in this context, we limit ourselves to simple tilings) that is closed under the translation action α and topologically closed under the topology given by the previously defined tiling metric.*

We may define equivariant and A-equivariant maps in the same way as before, as continuous maps between tiling spaces that behave well under translations (up to a multiplication by the matrix A), and thus we may define factor maps and conjugacies in the same way as in shift spaces. However, while the tiling topology is a continuous analogue to the shift topology, it may be the case that an equivariant map cannot be described via a local function (as in, there is no direct analogue to the Curtis–Hedlund–Lyndon in the case of general tilings, although there are somewhat weaker results in this direction that apply with more generality), and thus it is important to note when it happens.

Definition 3.41 *Let Ω_1, Ω_2 be two tiling spaces. We say that Ω_1 and Ω_2 are* ***mutually locally derivable*** *or MLD (sometimes written $\Omega_1 \leftrightsquigarrow \Omega_2$) if there exists a conjugacy $f\colon \Omega_1 \to \Omega_2$ such that there exists a finite radius $R > 0$ such that, whenever $T_1, T_2 \in \Omega_1$ match in a ball of radius R around the origin, $f(T_1)$ and $f(T_2)$ match on a ball of radius 1 around the origin. Note that in this situation f^{-1} has the same property.*

Of course, MLD tiling spaces are conjugate, but the converse is false; see the book by Sadun [89] for some examples. The concept extends to a more general idea of **local derivability**, which refers to the existence of a local description of a function.

Similar to shift spaces, tiling spaces may be described in a variety of ways, including local rules in a similar way to shifts of finite type. An example of this is the Socolar-Taylor tiling, where the tiles are all the reflections and rotations by multiples of $\frac{1}{3}\pi$ of an hexagonal tile with lines of different colors as decoration; two tiles are allowed to be next to each other if and only if the corresponding lines meet in specific ways.

We shall be interested in the tiling analogue to substitutions as a source for examples. This produces tilings that are "self-similar", in the sense that if λT is the tiling obtained from T by expanding each tile of T by a factor of $\lambda > 1$, the tiling spaces obtained as orbit closures of T and λT are MLD. The situation we shall be mostly interested in is when, given a finite set of prototiles $\mathcal{T}$, there is a rule J that assigns to every $\tau \in \mathcal{T}$ a finite cluster $J(\tau)$ of tiles of $\mathcal{T}$ satisfying the rules of a simple tiling and whose union is $\lambda\tau$; we call this an **inflation rule**. By iterating this process, we may obtain tilings $J^k(\tau)$ that cover any set of the form $\lambda^k\tau$, which are obtained by applying J to every tile from $J^{k-1}(\tau)$ and translating appropriately; this also allows us to extend J to any cluster of tiles from $\mathcal{T}$. As these sets have nonempty interior, we thus have a way to cover any arbitrarily large ball with tiles from $\mathcal{T}$, in a way dictated by the inflation rule J.

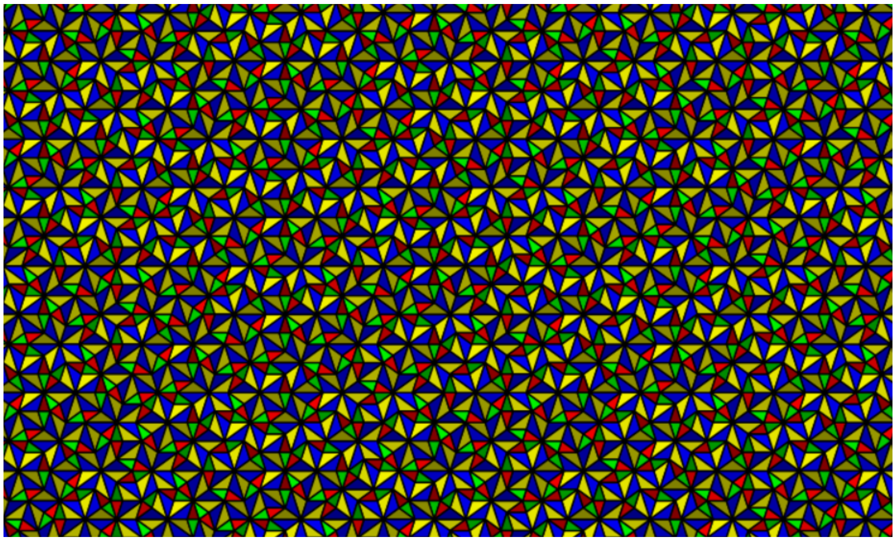


Figure 3.3: A section of the Penrose tiling, a well-known example of aperiodic inflation tiling.

Definition 3.42 *Given an inflation rule J, the corresponding **inflation tiling space** Ω_J is the set of all tilings T where every finite cluster of tilings from T is a translate of a subcluster of tiles of the cluster $J^k(\tau)$.*

Known examples include the chair and table tiling described by Olli [83], the half-hex tiling and the Penrose tiling. One-dimensional symbolic substitutions may be also seen as a specific subcase of inflations, even when they are not of constant length: for instance, the golden mean substitution $\theta\colon 0 \mapsto 01, 1 \mapsto 0$ may be seen as a one-dimensional inflation rule with two prototiles, the intervals $I_1 = [0,1]$ and $I_0 = [0,\varphi]$ (where φ is the golden mean), and an inflation factor of φ. The rule J transforms the interval I_1 into an interval of length φ, so that we may cover it with I_0, while $J(I_0)$ is an interval of length $\varphi^2 = \varphi + 1$, so we may tile it with a copy of I_0 followed with a copy of I_1. We note that, by checking the lengths of the tiles in a tiling from Ω_J, we obtain a sequence of symbols from X_θ and vice versa.

Note that, if we identify a tiling T with the point sets $\Lambda_\tau(T)$ indicating the positions of translates of the prototile τ, one may define finite sets $F_{\tau,\tau'} \Subset \mathbb{R}^d$ such that the cluster $J(\tau)$ contains a translation of $\tau' \in \mathcal{T}$ at $\boldsymbol{p}$, for every $\boldsymbol{p} \in F_{\tau,\tau'}$. Then, for any tiling $T \in \Omega_J$, the inflation rule J defines a new tiling $J(T) \in \Omega_J$ given by:

$$\Lambda_\tau(J(T)) = \bigcup_{\tau' \in \mathcal{T}} \lambda \cdot (\Lambda_{\tau'}(T) + F_{\tau',\tau}),$$

which is a useful form for computation. By ordering $\mathcal{T}$ in any way, we may define a square matrix with entries $[|F_{\tau,\tau'}|]_{\tau,\tau' \in \mathcal{T}}$, which serves the same purpose as the substitution matrix defined for symbolic substitutions. Thus, primitivity of this matrix accomplishes a similar role by ensuring minimality of the action α and, consequently, that there exists some J-periodic point T such that Ω_J is the orbit closure of T under translation.

Part II

Substitutive and hierarchical systems

Chapter 4

Extended symmetries in bijective substitutions and the Robinson tiling

The contents below correspond to the paper *Extended symmetry groups for multidimensional subshifts with hierarchical structure*, developed as part of the current thesis work. Some commentary and additional background has been added, while other sections have been simplified to prevent redundancy.

4.1. Introduction

Symbolic systems are a well known and thoroughly studied family of dynamical systems; among them, subshifts [62,71] take a central place due to the simplicity of the description of the shift action, with the complexity of a system manifesting itself in the description of the phase space instead. Thus, comparing different kinds of subshifts and determining whether they are "the same" in some sense (topological conjugacy, measure-theoretical isomorphism, shift equivalence, etc.) is a central problem in the theory of symbolic dynamics, followed by the search of mathematical objects that distinguish between "different" subshifts.

In what follows, X will denote a $\mathbb{Z}^d$-subshift. Among the "distinguishing objects" or **conjugacy invariants** the **automorphism group** $\mathrm{Aut}(X, \mathbb{Z}^d)$ is noteworthy, as it is comprised by self-conjugacies that preserve the dynamic structure of a subshift and, thus, the structure of the group $\mathrm{Aut}(X, \mathbb{Z}^d)$ contains important information about the nature of the dynamics of the subshift X. For instance, for a full $\mathbb{Z}$-shift $\mathcal{A}^{\mathbb{Z}}$, the automorphism group is "large" among countable groups [19] (by group-theoretic standards), having subgroups isomorphic to all finite groups, to $\bigoplus_{n=1}^{\infty} \mathbb{Z}$ and to the free group on two generators $\mathbf{F}_2$. This suggests that the shift action on the full $\mathbb{Z}$-shift exhibits many kinds of wildly different behaviors at once, such as periodic points, transitive orbits, an uncountable number of asymptotic pairs of points, etc., and such dynamical properties manifest in the increased complexity of the group $\mathrm{Aut}(\mathcal{A}^{\mathbb{Z}}, \mathbb{Z})$. By contrast, Sturmian subshifts X_α (where α refers to the angle of the associated rotation) have a trivial automorphism group $\mathrm{Aut}(\mathsf{X}_\alpha, \mathbb{Z}) = \langle \sigma \rangle \cong \mathbb{Z}$, which is consistent with our intuition as they are minimal subshifts with low complexity [29,30,33]. Here, any single point has enough information to determine the structure of the whole group. This suggests that strong structural constraints on the points of a subshift should make $\mathrm{Aut}(X, \mathbb{Z}^d)$ simple enough to be tractable. The book by Kitchens [62] compiles some classic results about automorphisms in subshifts, e.g. the aforementioned theorems by Boyle, Lind and Rudolph in [19]. Similarly, the work by Olli [83] deals with the computation of this group under strong rigidity conditions, with examples such as the two-dimensional chair tiling.

However, as by definition an automorphism is a translation-commuting mapping, when we deal with groups that have a rich geometrical structure (such as $\mathbb{Z}^d$, for $d > 1$) we see that the structure of the automorphism group is not sensitive to geometrical symmetries of a non-translational nature; this precludes $\mathrm{Aut}(X, \mathbb{Z}^d)$ from "detecting" differences in large-scale structures coming from such symmetries between two different subshifts. We can see some antecedents of the "geometric" phenomena that arise in higher-dimensional contexts in e.g. the algebraic systems studied by Kitchens and Schmidt [63, 90]. Thus, we would like to expand our scope to a larger group of homeomorphisms $f\colon X \to X$, that should satisfy identities on the lines of:

$$f \circ (\text{rigid symmetry}) = (\text{rigid symmetry}) \circ f,$$

as this allows for the homeomorphism f to "interact" properly with structural symmetries coming from rotations, reflections, shearing, etc. For the multidimensional groups $\mathbb{Z}^d, d > 1$, we may think of rigid symmetries as particular cases of affine transformations. In this context, if we think of shifts by elements of $\mathbb{Z}^d$ as geometric translations, we may attempt to formalize the desired equality above by employing the following property: a translation followed by an affine transformation equals the same affine transformation followed by a (possibly different) translation. In symbols, this results in an "almost shift-commuting" relation:

$$(\forall \boldsymbol{m} \in \mathbb{Z}^d)(\exists \boldsymbol{n} \in \mathbb{Z}^d)\colon f \circ \sigma_{\boldsymbol{m}} = \sigma_{\boldsymbol{n}} \circ f.$$

Algebraically, our new "generalized automorphisms" turn out to be exactly the elements of the **normalizer** group of $\langle \sigma \rangle = \{\sigma_{\boldsymbol{n}} : \boldsymbol{n} \in \mathbb{Z}^d\}$ in $\mathrm{Homeo}(X)$ (since $f^{-1} \circ \langle \sigma \rangle \circ f = \langle \sigma \rangle$), just in the same way as the group $\mathrm{Aut}(X, \mathbb{Z}^d)$ is the **centralizer** of $\langle \sigma \rangle$ in $\mathrm{Homeo}(X)$. Baake, Roberts and Yassawi study extensively these two groups and their relationship in the context of symbolic dynamics [13], although several precedents do exist (see e.g. the work by Goodson [47], the book on reversibility by O'Farrell and Short [82], or the study of the connection between flip-conjugacies and orbit equivalence, by Boyle and Tomiyama [18, 20]). As the relation between $\boldsymbol{n}$ and $\boldsymbol{m} \in \mathbb{Z}^d$ can be shown to be linear, i.e. $\boldsymbol{n} = A_f \boldsymbol{m}$ for some[1] $A_f \in \mathrm{GL}_d(\mathbb{Z})$, the general linear group $\mathrm{GL}_d(\mathbb{Z})$, representing the geometric structure of $\mathbb{Z}^d$, takes a prominent role in this context. Following [13], we call this normalizer the **extended symmetry group** of the $\mathbb{Z}^d$-subshift X, $\mathrm{Sym}(X, \mathbb{Z}^d)$.

In this work, we find strong restrictions for $\mathrm{Aut}(X, \mathbb{Z}^d)$ and $\mathrm{Sym}(X, \mathbb{Z}^d)$ for several $\mathbb{Z}^d$-subshifts which exhibit a hierarchical structure. We focus mainly on bijective substitutions, later diverting our attention towards the Robinson tiling. We prove the following results, all in the context of topological dynamics:

- For a bijective d-dimensional substitution θ, under reasonable assumptions, we show (Theorem 4.3) that the only nontrivial elements of $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$, modulo a shift, are relabeling maps, i.e. sliding block codes induced by a permutation of the alphabet $\mathcal{A}$, and thus this group is virtually-$\mathbb{Z}^d$, as it is the direct product of $\mathbb{Z}^d$ with a subgroup

[1] For a general group, a similar relationship holds, with $\mathrm{Aut}(G)$ taking the role of $\mathrm{GL}_d(\mathbb{Z})$. However, we shall not deal with this case in what follows.

of $S_{|\mathcal{A}|}$, the symmetric group over $\mathcal{A}$. This is a generalization of a result by Coven [28], and also a continuation of previous work such as the characterization by Lemańczyk and Mentzen [70] in a measurable, one-dimensional context, and later works in the subject [41, 57].

- In a similar fashion, we show (Theorem 4.6) that $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)$ is a finite extension of $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ by some subgroup of the group Q_d of symmetries of the d-dimensional cube. It is thus isomorphic to a subgroup of the product $(\mathbb{Z}^d \rtimes_\varphi Q_d) \times S_{|\mathcal{A}|}$, where the group action $Q_d \overset{\varphi}{\curvearrowright} \mathbb{Z}^d$ is the natural one coming from the identification of Q_d with a subgroup of $\mathrm{GL}_d(\mathbb{Z})$. An application of this result to the Thue–Morse d-dimensional substitution can be found as Example 4.3 below.
- With an argument of a similar nature, we compute the extended symmetry group of the Robinson tiling (and its minimal subshift), showing that it is a semidirect product of the corresponding automorphism group, which is known to be $\mathbb{Z}^2$ [34], with $Q_2 = D_4$, the group of symmetries of the square. These results are shown below as Proposition 4.11 and the following Corollary 4.13.

With those results, we showcase different proof techniques based on geometric considerations, highlighting how subshifts can be recognized as geometrical objects as well as dynamical ones.

4.2. Bijective substitutions and Coven's theorem

While this work is ultimately concerned with extended symmetries, it is helpful to have an explicit description of the automorphism group of the shift spaces under study first. We shall focus on **bijective** substitutions, defined before as those whose columns are all bijections.

Notation As previously, we reserve the letter d for the rank or dimension of the underlying group $\mathbb{Z}^d$, and the letter $\boldsymbol{s} = (s_1, \dots, s_d) \in \mathbb{Z}^d$ for a fixed "size" number, to be detailed below. As in Chapter 3, the symbol $\mathcal{A}$ will always denote the **alphabet**, whose elements are, consequently, called **symbols**.

Remark We may as well consider as a part of our study the analysis of what we could call the "**extended substitutive subshift**" X_θ^*, which would be the σ-orbit closure of the (finite) set of periodic points of θ_∞. This subshift is usually strictly larger than X_θ and consequently non-minimal, although with closely related dynamics. Some proofs below can be simplified when dealing with X_θ^*, due to the existence of "illegal" points with local properties preserved by automorphisms.

Given a pattern P over the alphabet $\{0, 1\}$, we write $\overline{P}$ for the pattern with the same support obtained by replacing all 1s by 0s and vice versa. It is easy to see that a substitution over the alphabet $\{0, 1\}$ is bijective if and only if $\theta(1) = \overline{\theta(0)}$. In this case, for any pattern P, $\theta(\overline{P}) = \overline{\theta(P)}$, and in particular $\theta^k(1) = \overline{\theta^k(0)}$. We can also define $\delta\colon \mathsf{X}_\theta \to \mathsf{X}_\theta$ by $\delta(x)_{\boldsymbol{k}} := \overline{x}_{\boldsymbol{k}}$. It is easy to see that δ is always a nontrivial automorphism of X_θ. Coven's result in one dimension [28] states that, up to a shift, δ is the only such automorphism:

Theorem 4.1 (Coven) *Let $\mathcal{A} = \{0, 1\}$ be a two-symbol alphabet. If $\theta\colon \mathcal{A} \to \mathcal{A}^\ell$ is a nontrivial, bijective, primitive substitution of constant length $\ell > 1$, then* $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}) \cong \mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z})$*, with* *every automorphism being of the form σ^n or $\delta \circ \sigma^n$, where $\sigma = \sigma_1$ is the elementary shift action and δ the aforementioned flip map.*

In larger alphabets there is a similar characterization. While we are concerned with the topological viewpoint only, regardless an important precedent worth mentioning is the characterization by Lemańczyk and Mentzen [70] of the nontrivial automorphisms of a bijective substitution, which (modulo a shift) are biunivocally associated to alphabet permutations that commute with the substitution θ; more precisely, they proved that:

$$\mathrm{Aut}(\mathsf{X}_\theta, \sigma)/\langle\sigma\rangle \cong \mathrm{Aut}(\mathsf{X}_\theta, \theta),$$

where the notations $\mathrm{Aut}(\mathsf{X}_\theta, \sigma)$ and $\mathrm{Aut}(\mathsf{X}_\theta, \theta)$ are used to distinguish the corresponding (semi-)group actions. Even though the method used in that work is oriented to the measurable case[2], certain ideas follow a similar pattern as the proof exhibited below for the multidimensional case, employing desubstitution (i.e., the property of recognizability from Lemma 3.20) in order to decompose a point from the shift as a concatenation of words $\theta^m(a)$ for any desired $m \geq 1$, which is then chosen appropiately depending on the automorphism under scrutiny. Our goal in what follows is to show that these results from Coven, Lemańczyk and Mentzen translate readily to the higher-dimensional case, in the context of topological dynamics.

The main step in the proof of our generalization of the above results lies in the following lemma[3]:

Lemma 4.2 *Let $\theta\colon \mathcal{A} \to \mathcal{A}^S$ be a bijective substitution with nontrivial support $S = [\mathbf{0}, \boldsymbol{s} - \mathbb{1}]$ over an alphabet $\mathcal{A}$, and suppose $f \in \mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ is an automorphism. Then, for any $x \in \mathsf{X}_\theta$ there exist $\boldsymbol{k}, \boldsymbol{\ell} \in \mathbb{Z}^d$ and a sufficiently large $m \geq 1$ such that both x and $f(x)$ are concatenations of patterns of the form $\theta^m(a), a \in \mathcal{A}$ arranged over a translation of a "grid" $\boldsymbol{s}^m \cdot \mathbb{Z}^d$, and such that the pattern with support $\boldsymbol{k} + \boldsymbol{p} + S^{(m)}$ (with $\boldsymbol{p} \in \boldsymbol{s}^m \cdot \mathbb{Z}^d$) in the grid corresponding to x determines uniquely the pattern with support $\boldsymbol{\ell} + \boldsymbol{p} + S^{(m)}$ in the grid corresponding to $f(x)$.*

Proof. As above, it is a direct consequence of Lemma 3.20 that for a fixed $m \geq 1$ any point $x \in \mathsf{X}_\theta$ is a concatenation of patterns of the form $\theta^m(a), a \in \mathcal{A}$ over a grid given by a translation of $\boldsymbol{s}^m \cdot \mathbb{Z}^d$. So we actually are proving the correspondence between these patterns in x and $f(x)$.

By its nature as a sliding block code, any automorphism $f \in \mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ can be assumed to be of radius $r \in \mathbb{N}_0$, i.e. to have a local function with window $[-r\mathbb{1}, r\mathbb{1}]$. Thus, for any subset $R \subseteq \mathbb{Z}^d$, $x|_R$ determines uniquely the configuration $f(x)|_{R^{\circ r}}$. From now on, f and r will be fixed. Consider then the support $S^{(m)}$ of θ^m. As S was deemed nontrivial, $S^{(m)}$ must be a d-dimensional rectangle of edge length at least 2^m in any direction, and thus for sufficiently large m (say, $m > \log_2(2r+1)$) the set $(S^{(m)})^{\circ r}$ is nonempty and a d-dimensional rectangle of edge length at least $2^m - 2r$ in all directions.

By Lemma 3.20, there are vectors $\boldsymbol{k}, \boldsymbol{\ell} \in \mathbb{Z}^d$ such that, for any $\boldsymbol{p} \in \boldsymbol{s}^m \cdot \mathbb{Z}^d$, $x|_{\boldsymbol{k}+\boldsymbol{p}+S^{(m)}}$ and $f(x)|_{\boldsymbol{\ell}+\boldsymbol{p}+S^{(m)}}$ are patterns of the form $\theta^m(a)$ for some $a \in \mathcal{A}$. We shall refer to these

[2] Note that in the measurable setting it is not necessary to distinguish X_θ from X_θ^*, as $\mathsf{X}_\theta^* \setminus \mathsf{X}_\theta$ has measure zero for the standard measures in this context, e.g. the frequency measure.

[3] We note that the proof below translates with no changes to X_θ^*.

rectangles as $K_{\boldsymbol{p}} := \boldsymbol{k} + \boldsymbol{p} + S^{(m)}$ and $L_{\boldsymbol{p}} := \boldsymbol{\ell} + \boldsymbol{p} + S^{(m)}$, respectively, for any $\boldsymbol{p} \in \boldsymbol{s}^m \cdot \mathbb{Z}^d$. Note that, since $S^{(m)} = [\boldsymbol{0}, \boldsymbol{s}^m - \mathbb{1}]$ is a set of representatives for $\mathbb{Z}^d / (\boldsymbol{s}^m \cdot \mathbb{Z}^d)$, the rectangles $K_{\boldsymbol{p}}$, indexed by all $\boldsymbol{p} \in \boldsymbol{s}^m \cdot \mathbb{Z}^d$, cover $\mathbb{Z}^d$ completely (and thus the $L_{\boldsymbol{p}}$ rectangles do so as well). Since we may replace $\boldsymbol{k}$ by any $\boldsymbol{k} + \boldsymbol{s}^m \cdot \boldsymbol{k}'$ (as then the new $K'_{\boldsymbol{p}}$ is just the old $K_{\boldsymbol{p}+\boldsymbol{k}'}$), we may choose $\boldsymbol{k}$ in a suitable way such that, for any $\boldsymbol{p} \in \boldsymbol{s}^m \cdot \mathbb{Z}^d$, $K_{\boldsymbol{p}}^{\circ r}$ has nonempty intersection with $L_{\boldsymbol{p}}$, say $I_{\boldsymbol{p}} := K_{\boldsymbol{p}}^{\circ r} \cap L_{\boldsymbol{p}}$. This is because the union of all $L_{\boldsymbol{p}}$ is the whole of $\mathbb{Z}^d$; we only need to note that, for a suitable choice of $\boldsymbol{k}$, the intersection $I_{\boldsymbol{0}} = K_{\boldsymbol{0}}^{\circ r} \cap L_{\boldsymbol{0}}$ is nonempty, and then use the fact that $K_{\boldsymbol{p}}$ and $L_{\boldsymbol{p}}$ are translations of $K_{\boldsymbol{0}}$ and $L_{\boldsymbol{0}}$ by the same vector. It is important to remark that, even though in most arguments we choose $\boldsymbol{k}$ and $\boldsymbol{\ell}$ from the set $S^{(m)} = [\boldsymbol{0}, \boldsymbol{s}^m - \mathbb{1}]$, as the obvious representatives of the cosets of $\boldsymbol{s}^m \cdot \mathbb{Z}^d$, it is not actually necessary to do so, and in particular in this proof $\boldsymbol{k}$ and $\boldsymbol{\ell}$ may be any two elements from $\mathbb{Z}^d$.

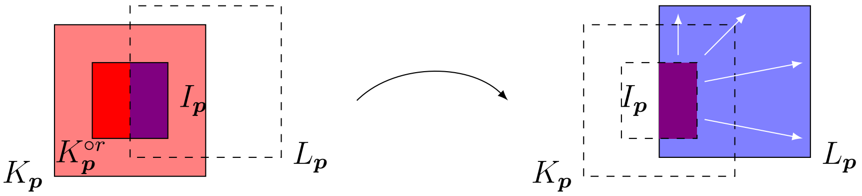


Figure 4.1: In the figure, we see how $x|_{K_{\boldsymbol{p}}} = \theta^m(a)$ (for some $a \in \mathcal{A}$) determines $f(x)|_{K_{\boldsymbol{p}}^{\circ r}}$ and, in particular, $f(x)|_{I_{\boldsymbol{p}}}$. Since the substitution is bijective, this forces $f(x)|_{L_{\boldsymbol{p}}}$ to equal $\theta^m(b)$ for some $b \in \mathcal{A}$ which depends solely on a.

As stated above, since θ (and thus θ^m) is a bijective substitution, then for any $a, b \in \mathcal{A}$ and any $\boldsymbol{q} \in S^{(m)}$ the condition $\theta^m(a)_{\boldsymbol{q}} = \theta^m(b)_{\boldsymbol{q}}$ implies $a = b$ and thus $\theta^m(a) = \theta^m(b)$. Because of this, the (unique) $b_{\boldsymbol{p}} \in \mathcal{A}$ such that the pattern $f(x)|_{L_{\boldsymbol{p}}}$ corresponds to (a translation of) $\theta^m(b_{\boldsymbol{p}})$ is entirely determined by the subpattern $f(x)|_{I_{\boldsymbol{p}}}$ (as $I_{\boldsymbol{p}}$ is nonempty), which in turn, as a subpattern of $f(x)|_{K_{\boldsymbol{p}}^{\circ r}}$, is entirely determined by $x|_{K_{\boldsymbol{p}}}$, which is of the form $\theta^m(a_{\boldsymbol{p}})$ for some $a_{\boldsymbol{p}} \in \mathcal{A}$ as well. Thus, for any $\boldsymbol{p} \in \boldsymbol{s}^m \cdot \mathbb{Z}^d$, $f(x)|_{L_{\boldsymbol{p}}}$ depends uniquely on $x|_{K_{\boldsymbol{p}}}$, as desired.

This proof shows that, associated to each f, there is a mapping $\tau_f \colon \mathcal{A} \to \mathcal{A}$ such that, for all $\boldsymbol{p} \in \boldsymbol{s}^m \cdot \mathbb{Z}^d$, $b_{\boldsymbol{p}} = \tau_f(a_{\boldsymbol{p}})$. This is enough to completely characterize f in terms of τ_f, and, using this, we can describe $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ explicitly:

Theorem 4.3 *For a nontrivial, bijective, primitive rectangular substitution θ, $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ is generated by the shifts and a finite set of relabeling maps of the form $\tau_\infty \colon \mathcal{A}^{\mathbb{Z}^d} \to \mathcal{A}^{\mathbb{Z}^d}$ given by permutations of the alphabet $\tau \colon \mathcal{A} \to \mathcal{A}$. Thus, $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ is isomorphic to the direct product of $\mathbb{Z}^d$ by some subgroup of $S_{|\mathcal{A}|}$. In particular, on the alphabet $\mathcal{A} = \{0, 1\}$, $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ is generated by the shifts and the relabeling map (flip map) $\delta(x) := \overline{x}$, and thus is isomorphic to $\mathbb{Z}^d \times (\mathbb{Z}/2\mathbb{Z})$.*

Proof. As remarked above, the relationship between $f(x)|_{L_{\boldsymbol{p}}}$ and $x|_{K_{\boldsymbol{p}}}$ can be stated as the existence of a mapping $\tau \colon \mathcal{A} \to \mathcal{A}$ (depending only on the automorphism f and perhaps on the chosen x) such that if $x|_{K_{\boldsymbol{p}}}$ is $\theta^m(a)$, then $f(x)|_{L_{\boldsymbol{p}}}$ is $\theta^m(\tau(a))$. The mapping τ does not depend on the chosen $\boldsymbol{p} \in \boldsymbol{s}^m \cdot \mathbb{Z}^d$ due to the CHL theorem.

Since f is an automorphism and thus invertible, we may apply the same observations to f^{-1}, obtaining another mapping of the alphabet $\eta \colon \mathcal{A} \to \mathcal{A}$. By choosing a sufficiently large m and appropiate values for $\boldsymbol{k}$ and $\boldsymbol{\ell}$, we see that $\eta \circ \tau(a) = a$ for all $a \in \mathcal{A}$. Since $|\mathcal{A}| < \infty$, τ

must be a bijection, as expected. In particular, for a binary alphabet $\mathcal{A} = \{0, 1\}$, τ is either the identity or the map $\tau(a) = 1 - a$.

Now, note that the m chosen for the proof of Lemma 4.2 can be replaced with any $m' > m$, with the argument remaining unchanged. Using desubstitution as given by Lemma 3.20, it is enough to prove the above equality for a point of the form $\theta^m(y)$. Taking $m' = m + 1$, the above characterization implies that:

$$\begin{aligned}(\exists \boldsymbol{k}, \boldsymbol{k'} \in \mathbb{Z}^d)(\exists \tau, \tau' \colon \mathcal{A} \to \mathcal{A}) \colon f(\theta_\infty^{m+1}(x)) &= \sigma_{\boldsymbol{k}}(\theta_\infty^m(\tau_\infty(\theta_\infty(x)))) \\ &= \sigma_{\boldsymbol{k'}}(\theta_\infty^{m+1}(\tau'_\infty(x))),\end{aligned}$$

and since $\boldsymbol{k} \equiv \boldsymbol{k'} \pmod{\boldsymbol{s}^m}$, this implies that each pattern $\theta^{m+1}(\tau'(a))$ with $a \in \mathcal{A}$ is a concatenation of the patterns $\theta^m(\tau(b))$, where the b are the corresponding symbols of the pattern $\theta(a)$. But by definition $\theta^{m+1}(\tau'(a)) = \theta^m(\theta(\tau'(a)))$, and the mapping θ is injective; thus, $\theta(\tau'(a)) = \tau(\theta(a))$, i.e. the relabeling τ must send patterns of the form $\theta(b)$ to other patterns of the form $\theta(b')$.

By replacing θ with a suitable power, we may assume that for the bottom left corner $\mathbf{0}$ of the support S the equality $\theta(a)_{\mathbf{0}} = a$ holds. Thus, $\theta(\tau'(a))$ has $\tau'(a)$ in this position, while $\tau(\theta(a))$ has $\tau(a)$ in the same position, i.e. $\tau(a) = \tau'(a)$. As this applies to any symbol a, we conclude that $\tau = \tau'$ and that τ and θ commute, i.e. $\theta_\infty \circ \tau_\infty = \tau_\infty \circ \theta_\infty$ as mappings $\mathcal{A}^{\mathbb{Z}^d} \to \mathcal{A}^{\mathbb{Z}^d}$. Applying this result to the identity with f above, we conclude that:

$$(\exists \boldsymbol{k} \in \mathbb{Z}^d) \colon f(\theta_\infty^m(x)) = \sigma_{\boldsymbol{k}} \circ \tau_\infty(\theta_\infty^m(x)),$$

and since τ_∞ is an automorphism of the full shift, it naturally commutes with $\sigma_{\boldsymbol{k}}$, which gives an explicit form for f for points of the form $\theta_\infty^m(x)$. Any other point of X_θ is a shift of a point of this form, and thus the result holds; in particular, for a binary alphabet, when $\tau(x) = 1 - x$ necessarily its extension to the full shift is $\tau_\infty = \delta$.

Now, for any $f \in \mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ and any $x \in \mathsf{X}_\theta$, $f(x)$ is of the form $\theta_\infty \circ \sigma_{\boldsymbol{\ell}-\boldsymbol{k}}(x)$ for some bijection $\theta \colon \mathcal{A} \to \mathcal{A}$. Since f commutes with the shift action, the equality $f|_{\mathrm{Orb}_\sigma(x)} = (\tau_\infty \circ \sigma_{\boldsymbol{\ell}-\boldsymbol{k}})|_{\mathrm{Orb}_\sigma(x)}$ holds. The result then follows by minimality[4].

Remark Note that the proofs of Lemma 4.2 and the following results also provide a necessary condition for a bijection $\tau \colon \mathcal{A} \to \mathcal{A}$ to induce a relabeling map $\tau_\infty \in \mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$, namely, that $\theta_\infty \circ \tau_\infty = \tau_\infty \circ \theta_\infty$ (under the condition that $\theta(a)_{\mathbf{0}} = a$ for all $a \in \mathcal{A}$, replacing θ by a suitable power if needed). This is essentially equivalent to the condition from [70] in the measurable case, as given by the isomorphism $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)/\langle\sigma\rangle \cong \mathrm{Aut}(\mathsf{X}_\theta, \theta_\infty)$. By compactness, this condition is also sufficient, providing an explicit description of the group $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ in terms of the patterns $\theta(a), a \in \mathcal{A}$. This condition may be restated in terms of the **columns** of the substitution θ, which are the bijections $\theta_{\boldsymbol{k}} \colon \mathcal{A} \to \mathcal{A}$, $a \mapsto \theta(a)_{\boldsymbol{k}}$, for $\boldsymbol{k} \in S$; this leads to a complete characterization of relabeling maps, as seen in the work by Frank [41], or in e.g. [57] in the one-dimensional case. Taking this previous work into account, we actually show that there are no other automorphisms besides these relabeling maps, making the aforementioned results a complete description of $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$.

[4] For the extension X_θ^*, since we lack minimality, we need to use the fact that X_θ^* is a finite union of orbit closures, the decomposition $f = \sigma_{\boldsymbol{k}} \circ \tau_\infty$ applying to each of them. Since all these orbit closures must contain X_θ, the relabeling τ must be the same for all of them.

Example An example of a bijective substitution arising naturally from tiling theory is the symbolic representation of the **chair tiling**. This tiling of $\mathbb{R}^2$ by 2×1 rectangles has a natural symbolic representation as seen in Figure 4.2, as a substitution θ in a four-letter alphabet corresponding to arrows pointing up, down, left and right.

The second power θ^2 of this substitution satisfies the property that all four corners of $\theta^2(a)$ have the symbol a. Suppose that there is an automorphism that is not a shift. Then, by appropriately composing this automorphism with a shift map, we may assume that it is a relabelling map, i.e. a radius zero automorphism τ_∞, where τ is a bijection $\mathcal{A} \to \mathcal{A}$, unequal to the identity. Thus, τ must commute with θ^2, i.e. $\theta^2(\tau(a))_{\boldsymbol{k}} = \tau(\theta^2(a)_{\boldsymbol{k}})$ for every $\boldsymbol{k} \in \operatorname{supp}(\theta^2)$; however, from the fact that the symbol a appears in all four positions of some edge of $\theta^2(a)$, and this edge is distinct for each value of a, we can easily see that the only mapping with this property is $\mathrm{id}_{\mathcal{A}}$. Thus, there are no nontrivial automorphisms in this shift space. This answers a question by Olli [REF].

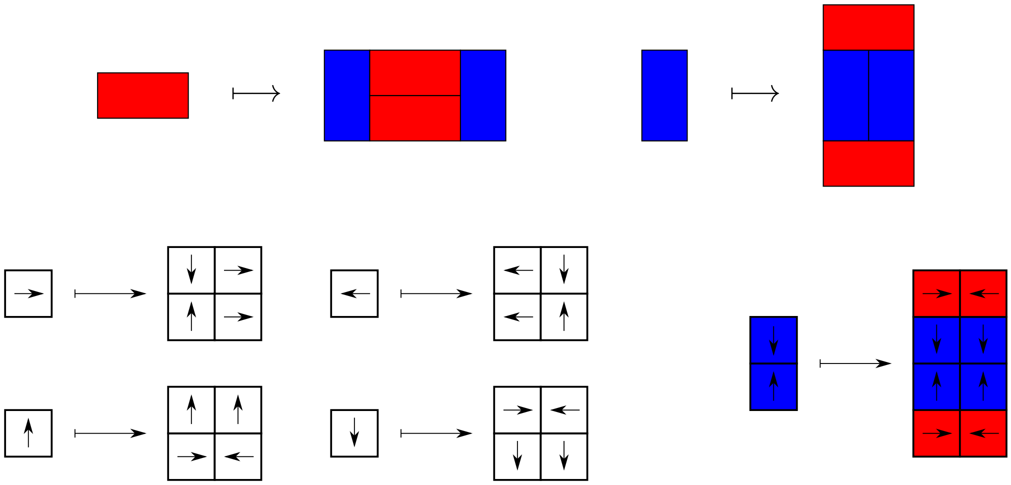


Figure 4.2: The inflation rule for the table tiling (above) and its symbolic representation as a substitution on a four-letter alphabet (below).

4.3. Extended symmetries of bijective substitutions

Our next goal is to obtain generalizations of the previous result in the domain of extended symmetries. These are a generalization of automorphisms, which introduce an additional degree of flexibility by allowing, besides the standard local transformations given by a sliding block code, to "deform" the underlying $\mathbb{Z}^d$ lattice, by rotation, reflection, shearing or other effects of a geometric nature. This additional degree of freedom is captured by a group automorphism of $\mathbb{Z}^d$, i.e. an element of $\mathrm{GL}_d(\mathbb{Z})$.

The basic premises of the theory of extended symmetries of subshifts may be studied in [13]. Remember that, as in Definition 3.9, an **extended symmetry** of a shift space X is an homeomorphism $f\colon X \to X$ which is A-equivariant for some invertible matrix $A \in \mathrm{GL}_d(\mathbb{Z})$, i.e.:

$$(\forall \boldsymbol{p} \in \mathbb{Z}^d)\colon f \circ \sigma_{\boldsymbol{p}} = \sigma_{A\boldsymbol{p}} \circ f,$$

and that the collection of all such functions is a group, $\mathrm{Sym}(X, \mathbb{Z}^d)$, containing the automorphism group $\mathrm{Aut}(X, \mathbb{Z}^d)$ as a normal subgroup. Under our standard hypothesis (namely, a faithful shift action) the matrix A_f associated to an extended symmetry f is uniquely determined and thus there is an obvious mapping $\psi\colon \operatorname{Sym}(X,\mathbb{Z}^d) \to \operatorname{GL}_d(\mathbb{Z}), f \mapsto A_f$. It is also easy to see that ψ is a group homomorphism:

$$f \circ g \circ \sigma_{\boldsymbol{p}} = f \circ \sigma_{\psi(g)\boldsymbol{p}} \circ g = \sigma_{\psi(f)(\psi(g)\boldsymbol{p})} \circ f \circ g,$$

consequently, $\psi(f \circ g) = \psi(f)\psi(g)$. Evidently, $\psi(f) = I_d$ (the identity matrix) if and only if f is a traditional automorphism of X, i.e. $\ker(\psi) = \operatorname{Aut}(X,\mathbb{Z}^d)$. This implies that the quotient group $\operatorname{Sym}(X,\mathbb{Z}^d)/\operatorname{Aut}(X,\mathbb{Z}^d)$ is isomorphic to a subgroup of $\operatorname{GL}_d(\mathbb{Z})$, and thus, determining the nature of the latter quotient in terms of $\operatorname{GL}_d(\mathbb{Z})$ is a very useful tool to describe $\operatorname{Sym}(X,\mathbb{Z}^d)$.

As is the case for automorphisms, this allows us to show that whenever two points match on a "large" set $R \subseteq \mathbb{Z}^d$, their images under an extended symmetry f match as well on a large set, which depends on f and R. More precisely, if we suppose w.l.o.g. that the support U (as defined in the theorem) of the symmetry f is of the form $[-r\mathbb{1}, r\mathbb{1}]$, then:

$$x|_R = y|_R \implies f(x)|_{\psi(f)[R^{\circ r}]} = f(y)|_{\psi(f)[R^{\circ r}]}.$$

In particular, if R is a half-space, the set $\psi(f)[R^{\circ r}]$ is a half-space as well.

Notation Given a d-tuple $\boldsymbol{u} = (u_1, \ldots, u_d) \in \{-1,+1\}^d$, we will denote by $Q_{\boldsymbol{u}}$ the **quadrant** $u_1\mathbb{N}_0 \times u_2\mathbb{N}_0 \times \cdots \times u_d\mathbb{N}_0$. Any translate $\boldsymbol{k} + Q_{\boldsymbol{u}}$ will also be called a quadrant (of vertex $\boldsymbol{k}$). Notice that $\mathbb{Z}^d$ can be written as the disjoint union of 2^d quadrants.

Our goal is to characterize the group $\operatorname{Sym}(X,\mathbb{Z}^d)/\operatorname{Aut}(X,\mathbb{Z}^d)$ when $X = \mathsf{X}_\theta$, θ being a nontrivial, bijective, primitive substitution, and then use this characterization to describe the extended symmetry group explicitly as a semidirect product, whenever possible. We shall see that X_θ has some distinguished points with **fractures**, that is, these points are comprised of a finite number of large subconfigurations (with a quadrant, a half-space, etc. as support) "glued together" in a somewhat independent way. Therefore, there are distinct points that match in a large subconfiguration (such as in a half-space) but are different outside the support of this configuration, in such a way that these differences may be "detected" locally. The latter implies, due to the generalized form of the CHL theorem, that extended symmetries have to preserve these points with fractures, in the sense that points that exhibit such a behavior are to be mapped to other points with similar characteristics. In particular, by analyzing these points adequately, we can deduce strong restrictions on the matrices $\psi(f) \in \operatorname{GL}_d(\mathbb{Z})$ for any $f \in \operatorname{Sym}(X,\mathbb{Z}^d)$, as the set of possible "shapes" of the fractures must be preserved by the matrix $\psi(f)$. The following lemma provides us with pairs of points that satisfy this general idea of "fractures":

Lemma 4.4 *Let θ be a nontrivial, bijective, primitive substitution such that X_θ has faithful $\mathbb{Z}^d$-shift action over the alphabet $\mathcal{A}$, and let $\boldsymbol{e}_i$ be any element of the canonical basis of $\mathbb{Z}^d$. Then there exist two points $x, y \in \mathsf{X}_\theta$ such that, for any $\boldsymbol{n} = n_1\boldsymbol{e}_1 + \ldots + n_d\boldsymbol{e}_d \in \mathbb{Z}^d$, $x_{\boldsymbol{n}} = y_{\boldsymbol{n}}$ if, and only if, $n_i \geq 0$; that is, x and y match exactly on a half-space that is a union of 2^{d-1} quadrants.*

Proof. For simplicity, we shall assume without loss of generality that $i = 1$. Since the action $\mathbb{Z}^d \overset{\sigma}{\curvearrowright} \mathsf{X}_\theta$ is faithful and minimal, there must be symbols $a, b, c \in \mathcal{A}$, with $b \neq c$, such that, for some points $x, y \in \mathsf{X}_\theta$, $x_{\mathbf{0}} = y_{\mathbf{0}} = a$ but $x_{-\boldsymbol{e}_1} = b, y_{-\boldsymbol{e}_1} = c$. If this were not the case, for any point $x \in \mathsf{X}_\theta$ the symbol $x_{\boldsymbol{k}}$ would determine $x_{\boldsymbol{k}+\boldsymbol{e}_1}$ uniquely. Since $|\mathcal{A}| < \infty$ this would result in a direction of periodicity shared by all points in X_θ, contradicting the faithfulness of the action.

As usual, we may replace θ by θ^m for a sufficiently large m such that every periodic point of θ is a fixed point of θ^m. By the previous observation, there exist two fixed points $x', y' \in \mathsf{X}_\theta$ such that $x'_{\mathbf{0}} = y'_{\mathbf{0}} = a$ and $x'_{-\boldsymbol{e}_1} = b, y'_{-\boldsymbol{e}_1} = c$. Those are obtained by iterating the substitution over the points x, y from the previous paragraph and taking a convergent subsequence, which exists by compactness. Since x' and y' are fixed points of the substitution, these symbols determine the corresponding quadrants entirely, and thus x' and y' match on the subset $Q_{\mathbb{1}} = \mathbb{N}_0^d$ but (due to bijectiveness) differ in every symbol from $(-\mathbb{N}) \times \mathbb{N}_0^{d-1}$.

Now, take the direction $\boldsymbol{p} = (0, -1, -1, \dots, -1)$. The pair of points $\sigma^m_{k\boldsymbol{p}}(x')$ and $\sigma^m_{k\boldsymbol{p}}(y')$ match on the set $E_k := \{n_1\boldsymbol{e}_1 + \ldots + n_d\boldsymbol{e}_d : n_1 \geq 0, n_2, \dots, n_d \geq -k\}$, and differ on every position in the set F_k defined by the same inequalities except for $n_1 < 0$. We may take a common convergent subsequence of $(\sigma^m_{k\boldsymbol{p}}(x'))_{k\geq 0}$ and $(\sigma^m_{k\boldsymbol{p}}(y'))_{k\geq 0}$, converging, respectively, to a pair of points x^* and y^* (note that such a pair exists by compactness). Due to the nature of convergence in shift spaces, x^* and y^* must match in E_k for infinitely many values of k, and thus match in $\bigcup_{k\in\mathbb{N}} E_k = \mathbb{N}_0 \times \mathbb{Z}^{d-1}$, and, simultaneously, they must differ in every position in the set $\bigcup_{k\in\mathbb{N}} F_k = (-\mathbb{N}) \times \mathbb{Z}^{d-1}$, as desired.

Remark In the extension X^*_θ, the existence of these kinds of points is easier to see, as they can be constructed directly from appropiate pairs of seeds. In fact, we can go further and create pairs of points which match everywhere but in a specific quadrant, as seen in Figure 4.3. As stated before, such illegal points are "discarded" in the measurable setting; this hints that we may use the additional structure of X^*_θ in the topological case to gain insight on the actual subshift of interest, X_θ.

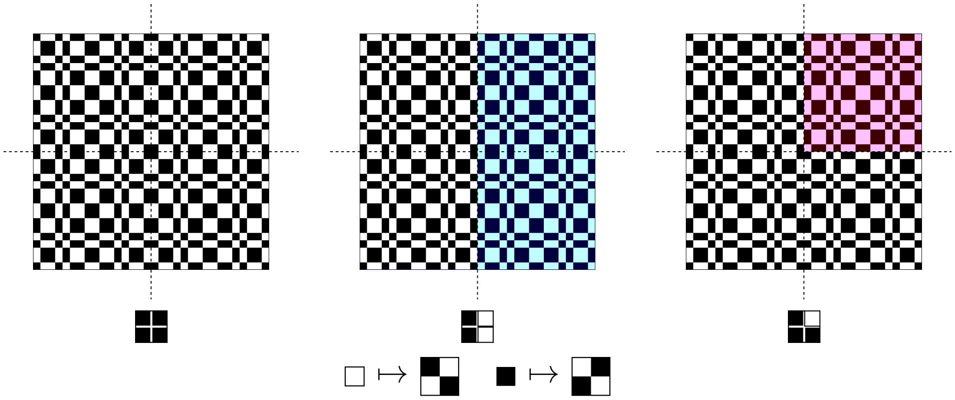


Figure 4.3: Points from the two-dimensional Thue-Morse substitution. The first two configurations correspond to (the central pattern of) two points $x, y \in \mathsf{X}_{\theta_{\mathrm{TM}}}$ matching exactly in one half-plane, as in Lemma 4.4. The third configuration is an "illegal" point $z \in \mathsf{X}^*_{\theta_{\mathrm{TM}}} \setminus \mathsf{X}_{\theta_{\mathrm{TM}}}$ from the extended substitutive subshift. The associated seeds and substitution rule are shown below.

To continue, we introduce some terminology. Remember that a (real) **hyperplane** of $\mathbb{R}^d$ is a $(d-1)$-dimensional affine subspace of $\mathbb{R}^d$; we shall call the intersection of such a hyperplane with $\mathbb{Z}^d$ a **discrete hyperplane**, as long as it satisfies the additional restriction of being a coset of a rank $d-1$ subgroup of $\mathbb{Z}^d$ (to avoid degenerate cases, e.g., a line with irrational slope can intersect $\mathbb{Z}^2$ in at most one point). Note that a discrete hyperplane H that passes through the origin is a rank $d-1$ direct summand of $\mathbb{Z}^d$, that is, there is some $\boldsymbol{v} \in \mathbb{Z}^d$ such that $\mathbb{Z}^d = H \oplus \mathbb{Z}\boldsymbol{v}$. Any discrete hyperplane can be written as:

$$H_{\boldsymbol{v}} := \{\boldsymbol{w} \in \mathbb{Z}^d : \langle \boldsymbol{v}, \boldsymbol{w} \rangle = 0\}, \text{ for some } \boldsymbol{v} \in \mathbb{Z}^d.$$

Note that the image of such a hyperplane under a matrix $A \in \mathrm{GL}_d(\mathbb{Z})$ is another such hyperplane, also given by some vector $\boldsymbol{v}'$ with integer coefficients. Such a hyperplane also defines two disjoint half-spaces $S_{\boldsymbol{v}}^+, S_{\boldsymbol{v}}^-$ (given by the inequalities $\langle \boldsymbol{v}, \boldsymbol{w} \rangle > 0$ and $\langle \boldsymbol{v}, \boldsymbol{w} \rangle < 0$, respectively) which together with $H_{\boldsymbol{v}}$ cover the whole of $\mathbb{Z}^d$. Any other half-space is a translation of a half-space of this type.

By the generalized CHL theorem, we verify that, given a pair of points x, y that match along a half-space $S_{\boldsymbol{e}_j}^{\pm}$ (as given by Lemma 4.4), their images under a fixed extended symmetry f match along a "large" set of the form $\psi(f)[(S_{\boldsymbol{e}_j}^{\pm})^{or}]$ as well. We shall use this to prove that, unless the image of a half-space $S_{\boldsymbol{e}_j}^{\pm}$ by $\psi(f)$ is itself a half-space of the form $S_{\boldsymbol{e}_{j'}}^{\pm}$ (where it may be the case that $j' \neq j$), the restriction of x to $S_{\boldsymbol{e}_j}^{\pm}$ determines $f(x)$ not only in $\psi(f)[(S_{\boldsymbol{e}_j}^{\pm})^{or}]$, but in $\psi(f)[(S_{\boldsymbol{e}_j}^{\mp})^{or}]$ (which is a translate of the complement of the previous set) as well, and from this we later infer that $x|_{S_{\boldsymbol{e}_j}^{\pm}}$ determines $f(x)$ in the whole plane. Thus, the existence of two distinct points that match in $S_{\boldsymbol{e}_j}^{\pm}$ contradicts the bijectivity of this hypothetical $f \in \mathrm{Sym}(X, \mathbb{Z}^d)$.

The previous definitions allow us to state a simple yet important property of the "grid of rectangles" that any point $x \in \mathsf{X}_\theta$ determines via the desubstitution property, as follows:

Lemma 4.5 *Let $\boldsymbol{n} \in \mathbb{Z}^d$ be a vector with positive entries, satisfying the condition $n_i > 1$ for all $1 \leq i \leq d$, and let S, S' be two disjoint half-spaces in $\mathbb{Z}^d$, given by vectors that are not (multiples of) the elements from the canonical basis. Then, for any $\boldsymbol{p} \in \mathbb{Z}^d$ and any sufficiently large $j \in \mathbb{Z}^+$ there is a $\boldsymbol{q} \in \boldsymbol{p} + \boldsymbol{n}^j \cdot \mathbb{Z}^d$ such that $\boldsymbol{q} + [\boldsymbol{0}, \boldsymbol{n}^j - \mathbb{1}]$ intersects both S and S'.*

Proof. First of all, note that the condition $S \cap S' = \varnothing$ implies that for some $\boldsymbol{v}$ (which is not a multiple of an element of the canonical basis) there exist $\boldsymbol{t}_1, \boldsymbol{t}_2 \in \mathbb{Z}^d$ such that $S = \boldsymbol{t}_1 + S_{\boldsymbol{v}}^+, S' = \boldsymbol{t}_2 + S_{\boldsymbol{v}}^-$. This is a direct consequence of the characterization of a half-plane via inequalities. We also note that it is enough to prove this result for $d = 2$, as we may restrict ourselves to a subspace $\mathbb{Z}\boldsymbol{e}_i \oplus \mathbb{Z}\boldsymbol{e}_j$ where $v_i \neq 0$ and $v_j \neq 0$ (such $i \neq j$ do exist because of the restriction on $\boldsymbol{v}$), and we also may assume $\boldsymbol{p} = \boldsymbol{0}$ by applying an adequate translation.

Note that, if we replace S and S' by subsets $S'' \subseteq S, S''' \subseteq S'$, proving the result for S'' and S''' does so for S and S' as well. Thus, without loss of generality, we may suppose that there exists some $\boldsymbol{v} = (v_1, v_2) \in \mathbb{Z}^2$ with $v_1 v_2 \neq 0$ and (for simplicity) $\gcd(v_1, v_2) = 1$, and some $C > 0$, such that $S = \{\boldsymbol{w} \in \mathbb{Z}^2 : \langle \boldsymbol{v}, \boldsymbol{w} \rangle \geq C\}$ and $S' = \{\boldsymbol{w} \in \mathbb{Z}^2 : \langle \boldsymbol{v}, \boldsymbol{w} \rangle \leq -C\} = -S$. Thus the following equality holds:

$$\mathbb{Z}^2 \setminus (S \cup S') = \{\boldsymbol{w} \in \mathbb{Z}^2 : |\langle \boldsymbol{v}, \boldsymbol{w} \rangle| \leq C\}.$$

This implies that $\mathbb{Z}^2 \setminus (S \cup S')$ is a disjoint union of translates of the discrete hyperplane (in the current context, a discrete line) $H_{\boldsymbol{v}} = \{\boldsymbol{w} \in \mathbb{Z}^2 : \langle \boldsymbol{v}, \boldsymbol{w} \rangle = 0\}$, and the latter equals $\mathbb{Z}\boldsymbol{u}$, where $\boldsymbol{u} = (-v_2, v_1)$; we shall suppose without loss of generality that both entries of $\boldsymbol{u}$ are positive.

Under the latter additional hypothesis, we see that for any $\boldsymbol{w} \in \mathbb{Z}^2$, the inner product $\langle \boldsymbol{v}, \boldsymbol{w} + \boldsymbol{e}_1 \rangle$ is strictly greater than $\langle \boldsymbol{v}, \boldsymbol{w} \rangle$ (with a difference of at least 1 since all quantities involved are integers). Iterating this, it is easy to see that any vectors from the line $C\boldsymbol{e}_1 + H_{\boldsymbol{v}}$ have inner product with $\boldsymbol{w}$ greater than C, and thus $C\boldsymbol{e}_1 + H_{\boldsymbol{v}} \subset S$. Similarly, $\langle \boldsymbol{v}, \boldsymbol{w} + \boldsymbol{e}_2 \rangle < \langle \boldsymbol{v}, \boldsymbol{w} \rangle$, which implies that $C\boldsymbol{e}_2 + H_{\boldsymbol{v}} \subset S'$ under the same reasoning. In particular, $C\boldsymbol{e}_1 \in S, C\boldsymbol{e}_2 \in S'$.

To conclude, note that the restriction $n_1, n_2 > 1$ implies that the rectangle $[\boldsymbol{0}, \boldsymbol{n}^j - \mathbb{1}]$ has at least 2^j elements on each side. Hence, for any sufficiently large j (say, $j > \log_2(C)$) this rectangle contains both $C\boldsymbol{e}_1$ and $C\boldsymbol{e}_2$ and it thus intersects both S and S', as desired. The same argument holds without the hypothesis $v_1, -v_2 > 0$ by replacing $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ by $\pm\boldsymbol{e}_1, \pm\boldsymbol{e}_2$ with appropiate signs.

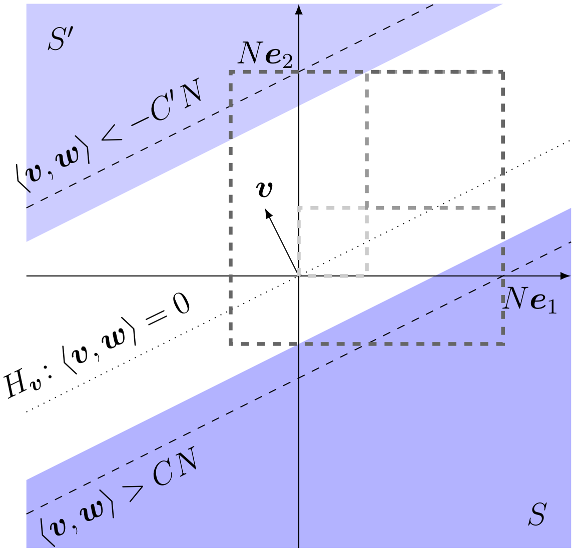

S'
$N\boldsymbol{e}_2$
$\langle \boldsymbol{v}, \boldsymbol{w} \rangle < -C'N$
$\boldsymbol{v}$
$N\boldsymbol{e}_1$
$H_{\boldsymbol{v}}$: $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = 0$
$\langle \boldsymbol{v}, \boldsymbol{w} \rangle > CN$
S

Figure 4.4: The situation in the proof of Lemma 4.5. As the side length of the rectangles associated with the substitution increases exponentially, the inner product $\langle \boldsymbol{v}, \boldsymbol{w} \rangle$ which determines whether $\boldsymbol{w}$ belongs to S or S' (or neither) takes sufficiently many different (integer) values inside any of these rectangles to ensure that at least one such rectangle intersects both S and S'.

Now, note that by shifting x and y by an appropiate sequence of vectors (chosen w.l.o.g. orthogonal to $\boldsymbol{v}$), we may ensure that the rectangle obtained by Lemma 4.5 is centered around the origin (e.g. if $\boldsymbol{n} = 2 \cdot \mathbb{1}$ then we may shift the corresponding points to ensure that the rectangle obtained has support $[-2^{m-1}, 2^{m-1} - 1]^d$). By compactness, and appealing to the continuity of the function that sends X_θ to its factor[5] $\mathbb{Z}_{\boldsymbol{n}}$ (which maps each point to the list of the vectors $\boldsymbol{k}$ corresponding to each shift in the representation from Lemma 3.20), we can replace both x and y by other two points x^* and y^* having the same property of matching in exactly one half-space and being different in every position in the other half-space, and

[5] Which is the maximal equicontinuous factor of X_θ in the height 1 case, but in this situation we do not actually need the maximality property.

also are given in such a way that a rectangle centered around the origin satisfies the property of Lemma 4.5 for infinitely many values of j.

This choice of points allows us to see that, given an extended symmetry $f\colon \mathsf{X}_\theta \to \mathsf{X}_\theta$, since x^* and y^* differ in a half-space S, $f(x^*)$ and $f(y^*)$ must differ in infinitely many positions in $\psi(f)[S]$ (which is a half-space as well, equalling $S_{\boldsymbol{v}}^+$ for some $\boldsymbol{v}$), and such discrepancies may be found at an arbitrarily large distance from the separating hyperplane $H_{\boldsymbol{v}}$; otherwise, by shifting in the direction of $\boldsymbol{v}$ we can construct by compactness two different points with the same image, contradicting the bijectiveness of $\boldsymbol{v}$. But any discrepancy between $f(x^*)$ and $f(y^*)$ is located in a sufficiently large rectangle of the kind given by Lemma 4.5 centered on the origin. This implies that, since the substitution is bijective, $f(x^*)$ and $f(y^*)$ must be different in all coordinates from $S_{\boldsymbol{v}}^+$ located inside this rectangle. Since the rectangles can be taken as large as desired, this implies that $f(x^*)$ and $f(y^*)$ must differ in every position from $S_{\boldsymbol{v}}^+$.

Now we may prove the core part of our characterization of the extended symmetries of X_θ, in the form of the following theorem[6]:

Theorem 4.6 *For a d-dimensional, nontrivial, bijective, primitive substitution θ, the group of all admissible lattice transformations of the subshift X_θ, $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)/\operatorname{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$, is isomorphic to a subset of the hyperoctahedral group[7] $Q_d \cong (\mathbb{Z}/2\mathbb{Z}) \wr S_d = (\mathbb{Z}/2\mathbb{Z})^d \rtimes S_d$, which represents the symmetries of the d-dimensional cube. Thus, the extended symmetry group $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)$ is virtually-$\mathbb{Z}^d$.*

Proof. To begin with, we need to show that, for an extended symmetry f, the matrix $\psi(f)$ must map half-spaces of the form $S_{\boldsymbol{e}_i}^\pm$ to (possibly different) half-spaces $S_{\boldsymbol{e}_j}^\pm$ of the same sort. Suppose that $\psi(f)[S_{\boldsymbol{e}_1}^+]$ is $S_{\boldsymbol{v}}^+$ for some $\boldsymbol{v}$ with at least two nonzero coordinates (that is, it is not a scalar multiple of a vector from the canonical basis). By Lemma 4.4, we know that there are two points $x, y \in \mathsf{X}_\theta$ such that $x_{\boldsymbol{n}} = y_{\boldsymbol{n}}$ if, and only if, $\langle \boldsymbol{n}, \boldsymbol{e}_1 \rangle \geq 0$. Without loss of generality, we assume that these two points have the properties described in the brief discussion after Lemma 4.5.

Note that the half-spaces $S_{\boldsymbol{e}_1}^-$ and $S_{\boldsymbol{e}_1}^+ - \boldsymbol{e}_1$ are a partition of $\mathbb{Z}^2$ and each of them determines $f(x)$ and $f(y)$ over the sets $\psi(f)[S_{\boldsymbol{e}_1}^-]^{\circ r}$ and $\psi(f)[S_{\boldsymbol{e}_1}^+ - \boldsymbol{e}_1]^{\circ r}$, which are translations of $S_{\boldsymbol{v}}^-$ and $S_{\boldsymbol{v}}^+$, respectively. In particular, by the definition of x and y, we have that, for some $C > 0$, $f(x)|_S = f(y)|_S$ but $f(x)|_{S'}$ and $f(y)|_{S'}$ differ at every coordinate, where $S = \{\boldsymbol{w} \in \mathbb{Z}^d : \langle \boldsymbol{v}, \boldsymbol{w} \rangle > C\}$ and $S' = \{\boldsymbol{w} \in \mathbb{Z}^d : \langle \boldsymbol{v}, \boldsymbol{w} \rangle < -C\}$.

By desubstitution, since x and y match in a very large set containing a quadrant, we see that for any $m > 0$ there exists some $\boldsymbol{k} \in \mathbb{Z}^d$ such that $f(x) = \sigma_{\boldsymbol{k}}(\theta^m(x')), f(y) = \sigma_{\boldsymbol{k}}(\theta^m(y'))$ for some $x', y' \in \mathsf{X}_\theta$. Thus, both $f(x)$ and $f(y)$ are concatenations of patterns of the form $\theta^m(a), a \in \mathcal{A}$ with support $\boldsymbol{p} + [\boldsymbol{0}, \boldsymbol{n}^m - \mathbb{1}]$, with $\boldsymbol{p} \in \boldsymbol{k} + \boldsymbol{n}^m \cdot \mathbb{Z}^d$. By Lemma 4.5, we can choose a sufficiently large m in order to ensure the existence of one such rectangle, say R, that has nonempty intersection with both S and S'.

[6] Once again, the same proof holds for $\mathrm{Sym}(\mathsf{X}_\theta^*, \mathbb{Z}^d)$, and so we omit explicit distinctions. However, for X_θ^* an alternative, simpler approach is possible, which is briefly discussed afterwards.

[7] This group, with operation $(n_1, \dots, n_d; \tau) * (m_1, \dots, m_d; \nu) = (n_1 + m_{\tau(1)}, \dots, n_d + m_{\tau(d)}; \tau\nu)$, is a particular case of a wreath product. A description of this group, its properties and relation to the geometry of the integer lattice $\mathbb{Z}^d$ can be found in [78], or in [2]. Wreath products in general are described in the book by James and Kerber [54], among others.

Since $f(x)|_S = f(y)|_S$, we must have $f(x)|_{S\cap R} = f(y)|_{S\cap R}$. As the substitution is bijective, this implies $f(x)|_R = f(y)|_R$ and, in particular, $f(x)|_{S'\cap R} = f(y)|_{S'\cap R}$. However, from our previous observation, we know that since x and y differ in every coordinate from $S^-_{\boldsymbol{e}_1}$, we must have $f(x)|_{S'\cap R} \neq f(y)|_{S'\cap R}$, which is a contradiction. As this contradiction arose from the hypothesis of $\boldsymbol{v}$ not being a scalar multiple of some element of the canonical basis (as in the hypothesis of Lemma 4.5), we must have $\boldsymbol{v}$ be some multiple of some $\boldsymbol{e}_i$. Since any positive multiple of $\boldsymbol{v}$ defines the same half-space up to a translation, we may assume $\boldsymbol{v} = \pm\boldsymbol{e}_j$ for some index $1 \leq j \leq d$. The same argument holds for any other element of the canonical basis besides $\boldsymbol{e}_1$.

Since the sequence of vectors $\psi(f)\boldsymbol{e}_i, 1 \leq i \leq d$, must consist of linearly independent members of the canonical basis, and thus corresponds to a permutation of this basis with added signs, the group $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)$ acts by permutation on the set of one-dimensional subspaces $\{\mathbb{Z}\boldsymbol{e}_1, \dots, \mathbb{Z}\boldsymbol{e}_d\}$. Thus, $\psi(f)$ must be given by a matrix that sends each $\boldsymbol{e}_i$ from the canonical basis to a vector $\pm\boldsymbol{e}_j$ and thus each column of $\psi(f)$ is such a vector. Since $\psi(f)$ is non-singular, it must be of the form:

$$\psi(f) = [(-1)^{t_1}\boldsymbol{e}_{\pi(1)} \mid (-1)^{t_2}\boldsymbol{e}_{\pi(2)} \mid \cdots \mid (-1)^{t_d}\boldsymbol{e}_{\pi(d)}],$$

where π is a permutation of $\{1, \dots, d\}$ and $t_1, \dots, t_d \in \{0, 1\}$. These matrices correspond to a finite subgroup of $\mathrm{GL}_d(\mathbb{Z})$ which is isomorphic to Q_d. Indeed, the set of all diagonal matrices of this form is isomorphic to $(\mathbb{Z}/2\mathbb{Z})^d$, while the set of all matrices with nonnegative entries of this form is isomorphic to S_d, and any matrix of the aforementioned form is a product of a permutation matrix with positive entries and a diagonal matrix in a unique way.

Thus, ψ can be seen as a group homomorphism $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d) \to Q_d$ by identifying the latter with the corresponding matrix group. Since $\ker(\psi) = \mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$, we conclude that $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)/\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d) \cong \mathrm{im}(\psi) \leq Q_d$, as desired.

Remark The proof given above is mostly combinatorial in nature, appealing to some geometric properties both inherent to the general structure of the subshifts involved and present in specific points of this space. We can take a different approach, closer to topological dynamics in the same vein as [13]; we shall proceed to give a sketch of this alternate method. As before, $\varphi\colon \mathsf{X}_\theta \twoheadrightarrow \mathbb{Z}_{\boldsymbol{n}}$ is the mapping from the substitutive subshift X_θ to its odometer factor.

Using arguments similar to the construction from Lemma 4.4 and the subsequent discussion, we can show that for any integer coordinate of a $\boldsymbol{m} \in \mathbb{Z}_{\boldsymbol{s}}$, we may insert a hyperplane in $\mathbb{Z}^d$, parallel to the corresponding coordinate hyperplane, in such a way that these hyperplanes induce a partition of $\mathbb{Z}^d$ into 2^r subsets $E_1, \dots, E_{2^r}$, each a finite union of 2^{d-r} quadrants, so that if $U \subset \mathbb{Z}^d$ is any set that intersects all the E_j, then any $x \in \mathsf{X}_\theta$ with $\varphi(x) = \boldsymbol{m}$ is entirely determined by $x|_U$. In particular, if $\varphi(x) = \boldsymbol{0}$, x is necessarily a periodic point of θ.

We can use this property to show that there exists an ℓ such that $|\varphi^{-1}[\boldsymbol{m}]| = \ell$ for any $\boldsymbol{m} \in \mathbb{Z}^d$, while $|\varphi^{-1}[\boldsymbol{m}]| \neq \ell$ for all $\boldsymbol{m} \in \mathbb{Z}_{\boldsymbol{s}} \setminus \mathbb{Z}^d$. Arguments akin to the ones shown in [13] show that extended symmetries preserve the cardinality of these fibers of the factor φ, hence they must map $\mathrm{Per}_\theta(\mathsf{X}_\theta)$ to itself. The result then follows by application of the extended CHL theorem.

Remark While Theorem 4.6 applies to both the standard substitutive subshift X_θ and the extended subshift X^*_θ, in the latter case we may use the properties of quadrants in $\mathbb{Z}^d$ and affine mappings for an alternate proof method. We summarize this alternative approach briefly, as we believe this proof scheme may apply to other kinds of subshifts with similar geometrical properties.

Via basic linear algebra and convexity arguments, we may show that the image of a quadrant Q under a matrix A cannot contain two disjoint quadrants; more precisely, as a quadrant is the set of all nonnegative integral linear combinations of some set of the form $\{\pm\boldsymbol{e}_1, \dots, \pm\boldsymbol{e}_d\}$, its image under an invertible matrix $A \in \mathrm{GL}_d(\mathbb{Z})$ is also the set of nonnegative linear combinations of some set of d vectors. As AQ is the intersection of a convex cone with $\mathbb{Z}^d$, if $\boldsymbol{p} \in \mathbb{Z}^d$ is a convex combination of two elements $\boldsymbol{q}, \boldsymbol{r} \in AQ$, then $\boldsymbol{p} \in AQ$ as well. Were AQ to contain two distinct quadrants, then, this allows us to show that AQ contains all of the (infinitely many) points with integral coordinates of some line in $\mathbb{R}^d$, and thus Q contains a line as well; however, such a line would eventually have a point where one of the coefficients of the associated linear combination is negative, a contradiction.

This results in a limitation of the "shape" of the cone AQ obtained as this image to specific configurations, which result on either AQ or $A^{-1}Q$ being strictly contained in a quadrant. This, in turn, forces some rectangles associated to the desubstitution box structure (from Lemma 3.20) to overlap both this cone and its complement in subsets of $\mathbb{Z}^d$ with arbitrarily large diameter. Hence, by the bijectiveness of θ, the configuration of symbols in this cone in $f(x)|_{AQ}$ is forced by the configuration $x|_{Q^c}$ outside of this quadrant.

Now, let x_1 and x_2 be two fixed points of θ (replacing it with a suitable power such that $\theta^k(a)$ has the symbol a on every corner, if needed), whose seed (central pattern) is equal in all but one symbol, in such a way as to ensure that those points have the same symbols in every quadrant except Q, in which they differ at every position. Then, as $A^{\pm 1}Q$ is entirely contained within some quadrant, the previous bijectiveness argument forces the images of x_1 and x_2 under $f^{\pm 1}$ to be the same. As f (and thus f^{-1}) is a bijection, this is absurd.

The previous result imposes a very strict limitation on the structure of the group $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)$; thus, with some additional information, we can compute this group explicitly. For instance:

Example The extended symmetry group of the subshift induced by the d-dimensional Thue–Morse substitution, given by:

$$\begin{aligned} \theta_{\mathrm{TM}} \colon \{0,1\} &\to \{0,1\}^{\{0,1\}^d} \\ a &\mapsto ((a + m_1 + \cdots + m_d) \text{ mód } 2)_{(m_1,\dots,m_d) \in \{0,1\}^d}, \end{aligned}$$

is a semidirect product of the form:

$$\mathrm{Sym}(\mathsf{X}_{\theta_{\mathrm{TM}}}, \mathbb{Z}^d) \cong (\mathbb{Z}^d \times \mathbb{Z}/2\mathbb{Z}) \rtimes Q_d,$$

generated by the shifts, the relabeling map $\delta(x) = \overline{x}$ and the $2^d d!$ rigid symmetries of the coordinate axes given by $(\varphi_A(x))_{\boldsymbol{n}} = x_{A\boldsymbol{n}}$, with $A \in Q_d$.

Indeed, by Theorem 4.6, $\mathrm{Sym}(\mathsf{X}_{\theta_{\mathrm{TM}}}, \mathbb{Z}^d)$ is a $(\mathbb{Z}^d \times (\mathbb{Z}/2\mathbb{Z}))$-by-$R$ group extension for some $R \leq Q_d$. Consider x to be the fixed point of θ_{TM} whose seed $x|_{\{-1,0\}^d}$ consists only of zeros. It is easy to verify that, for any $A \in Q_d$, $\varphi_A(x) = x$; since $\mathsf{X}_{\theta_{\mathrm{TM}}}$ is the orbit closure of x, this immediately shows that φ_A maps $\mathsf{X}_{\theta_{\mathrm{TM}}}$ to itself and thus φ_A is a valid extended symmetry. Hence, $R = Q_d$, which embeds into $\mathrm{Sym}(\mathsf{X}_{\theta_{\mathrm{TM}}}, \mathbb{Z}^d)$ by the map $\iota \colon A \mapsto \varphi_A$. Since $\psi(\varphi_A) = A$, ι is a right inverse for ψ and thus the extension splits, corresponding to the above mentioned semidirect product. With minor changes, we can show the same result for $\mathsf{X}^*_{\theta_{\mathrm{TM}}}$.

Example In a similar fashion, the symbolic version of the table tiling from Figure 4.2 may be immediately seen to have D_4 symmetry, as suggested by the corresponding inflation rule. We note that, in this case, the underlying local function of the extended symmetries must be non-trivial; for instance, for a reflection along the Y axis, this local function must preserve the vertical arrows but swap the right- and left-pointing arrows. This is a similar situation to the chair tiling example studied by Olli [83], whose extended symmetry group was described by Baake, Roberts and Yassawi [13].

Remark It is worth noting that, while having extended symmetries associated to the hyperoctahedral group may be the expected result due to the underlying box structure, the bijectiveness of θ plays an important role as well. For instance, consider the half-hex tiling, which is given by the inflation rule seen in Figure 4.5. Splitting each hexagon into three triangles, we see that there are 13 possible configurations of one triangle pointing upwards horizontally adjacent to one pointing downwards forming a rhombus.

Using this set of 13 rhombuses as an alphabet, the inflation rule becomes a non-bijective substitution θ which encodes the inflation. Indeed, we may retrieve a tiling consistent with this rule from every $x \in \mathsf{X}_\theta$ and vice versa; this is a similar situation to the chair and table tilings [83].

It is easy to see that applying an hexagonal symmetry to any tiling obtained via this inflation produces another tiling of the same thing. This fact may be used to produce a copy of D_6 inside of $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)/\operatorname{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ (it may be proved, even, that this quotient is exactly D_6), which is the symbolic encoding of the associated rotations and reflections. In particular, this quotient group has an element of order 6. As no matrix from D_4 has this order, we see that the bijectiveness hypothesis is essential in the proof of Theorem 4.6.

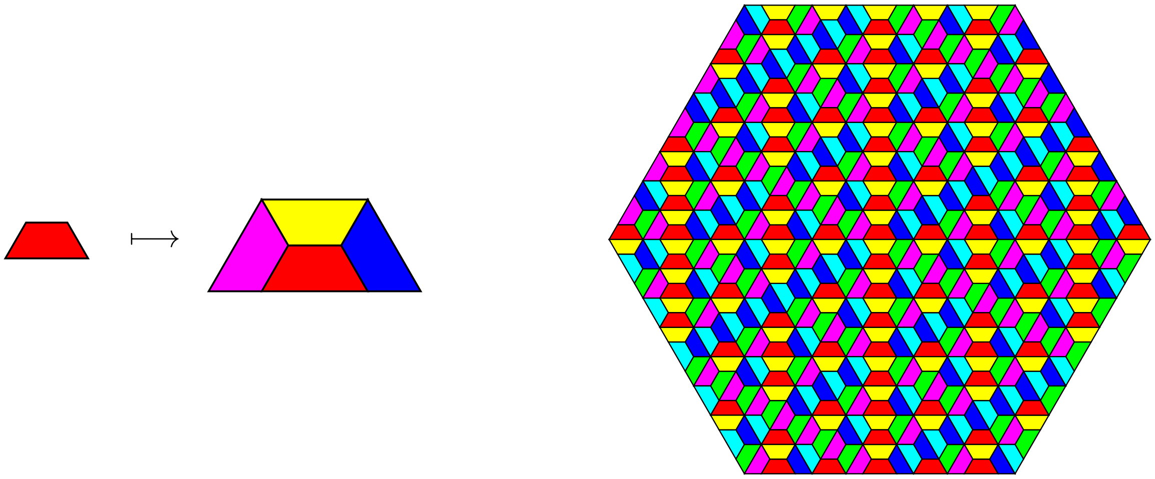


Figure 4.5: A half-hexagon and its image under the inflation rule. Rotations by multiples of $\frac{1}{3}\pi$ determine entirely this inflation rule.

4.4. The Robinson shift and fractures in subshifts

In the remaining section, we leave the setting of substitutive tilings from above and move on to analyze a well-known example of **strongly aperiodic** $\mathbb{Z}^2$-subshift, the Robinson shift [88].

Definition 4.7 *Let X be a $\mathbb{Z}^d$-subshift. We say X is **strongly aperiodic** if all points in X have trivial stabilizer, i.e., for all $x \in X$, $\sigma_{\boldsymbol{k}}(x) = x$ implies $\boldsymbol{k} = \boldsymbol{0}$.*

The Robinson shift is a two-dimensional nearest-neighbor shift with added local restrictions (and thus of finite type), whose alphabet consists of all the rotations and reflections of the five tiles from Figure 4.6, resulting in 28 different symbols.

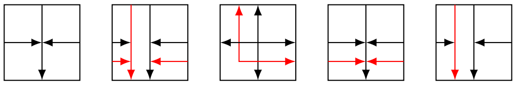


Figure 4.6: The five types of Robinson tiles, resulting in an alphabet of 28 symbols after applying all possible rotations and reflections. The third tile is usually called a **cross**.

The Robinson shift X_{Rob} is given by the following local rules:

(1) Every arrow head in a tile must be in contact with an arrow tail from an adjacent tile (nearest-neighbor rule). This is similar to the local rule of a Wang tiling (although not exactly equivalent; see [88] or [44] for details).

(2) There is a translation of the sublattice $2\mathbb{Z} \times 2\mathbb{Z}$ that only has rotations of the central tile of Figure 4.6 (which shall be referred to as a **cross**).

(3) Any other cross appears diagonally adjacent to one of the crosses from the sublattice of Rule (2). Namely, if the cross-only sublattice of a given point is $(2\mathbb{Z} \times 2\mathbb{Z}) + \boldsymbol{k}$, then any other cross is placed at one of the positions from $(2\mathbb{Z} \times 2\mathbb{Z}) + \boldsymbol{k} + \mathbb{1}$.

It is easy to see that those rules can be enforced with strictly local restrictions and thus X_{Rob} is a shift of finite type. These rules force the 28 basic tiles to form larger patterns with similar behavior to each of the five tiles (in particular, patterns of size $(2^n - 1) \times (2^n - 1)$ that behave as larger analogues of crosses and that are usually referred to as n-th order **supertiles**). By compactness, as we can always build larger supertiles from smaller ones, we can prove that X_{Rob} is a non-empty strongly aperiodic subshift. It is not minimal, but it has a unique minimal subsystem M_{Rob} (which is the factor of a subshift of finite type). Its automorphism group has previously been characterized, e.g. by Donoso and Sun [34], and the underlying behavior is similar to the one present in the chair tiling [83]:

Theorem 4.8 $\mathrm{Aut}(M_{\mathrm{Rob}}, \mathbb{Z}^2) = \langle \sigma_{(1,0)}, \sigma_{(0,1)} \rangle \cong \mathbb{Z}^2$.

From this result, it is possible to show that the same holds for X_{Rob}, namely that the only automorphisms of the Robinson shift are the trivial ones. We aim to extend this result by computing the extended symmetry group of the Robinson shift. For this, we need to introduce a distinguished subset of $\mathbb{Z}^2$ which represents part of the structure of a shift which is preserved by extended symmetries:

Definition 4.9 *Let X be a strongly aperiodic $\mathbb{Z}^2$-subshift. We say X has a **fracture** in the direction $\boldsymbol{q} \in \mathbb{Z}^2$ if there is a point $x^* \in X$, two disjoint half-planes $S^+, S^- \subseteq \mathbb{Z}^2$ separated*

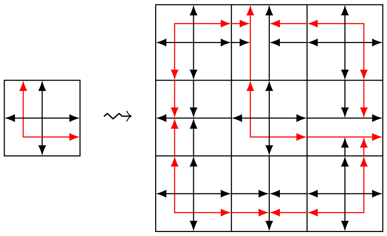


Figure 4.7: The formation of a second order supertile of size 3×3.

by $\mathbb{Z}\boldsymbol{q}$ *(i.e.* $S^+ \cap S^- = S^+ \cap \mathbb{Z}\boldsymbol{q} = S^- \cap \mathbb{Z}\boldsymbol{q} = \varnothing$*; it is not necessary that* $S^+ \cup S^- \cup \mathbb{Z}\boldsymbol{q} = \mathbb{Z}^2$*) and infinite different values* $k_1 < k_2 < k_3 < \ldots \in \mathbb{Z}$ *such that, for each* $j \in \mathbb{N}$*, there is a point* $x^{(j)} \in X$ *that satisfies the two conditions:*

$$x^{(j)}|_{S^+} = x^*|_{S^+}, \quad x^{(j)}|_{S^-} = \sigma_{k_j \boldsymbol{q}}(x^*)|_{S^-}.$$

Remark We exclude subshifts with periodic points from this definition as, if $x \in \operatorname{Per}_{\boldsymbol{p}}(X)$, we may take $k_j = j$ and $x^{(j)} = x$ for all values of j, resulting in a point with a fracture in the direction $\boldsymbol{p}$. This makes the definition of direction of fracture redundant with the concept of direction of periodicity, which is also preserved by extended symmetries.

Lemma 4.10 *Let* $\boldsymbol{q} \in \mathbb{Z}^2$ *be a direction of fracture for a two-dimensional strongly aperiodic subshift* X *and* $f \in \operatorname{Sym}(X, \mathbb{Z}^2)$*. Then* $\psi(f)\boldsymbol{q}$ *is a direction of fracture as well.*

Proof. Let $\boldsymbol{q}$ be a direction of fracture, and $x^*, (x^{(j)})_{j\in\mathbb{N}}, (k_j)_{j\in\mathbb{N}}$ be the associated points and magnitudes from the definition above. By the generalized CHL theorem, as $x^*|_{S^+} = x^{(j)}|_{S^+}$, then $f(x^*)|_{\psi(f)((S^+)^{\circ r})} = f(x^{(j)})|_{\psi(f)((S^+)^{\circ r})}$, where r is the radius of the symmetry f. By the same argument, and since $f \circ \sigma_{\boldsymbol{q}} = \sigma_{\psi(f)\boldsymbol{q}} \circ f$, we conclude that $f(x^{(j)})|_{\psi(f)((S^-)^{\circ r})} = \sigma_{k_j\psi(f)\boldsymbol{q}} \circ f(x^*)|_{\psi(f)((S^-)^{\circ r})}$.

Note that, since S^+ and S^- are half-planes disjoint from the linear subspace $\mathbb{Z}\boldsymbol{q}$, and $\psi(f)$ is a linear map, $(S^{\pm})^{\circ r}$ are also half-spaces and thus their corresponding images $\psi(f)((S^{\pm})^{\circ r})$ are half-spaces as well. As subsets of the images of disjoint sets, they are also disjoint from $\mathbb{Z}(\psi(f)\boldsymbol{q})$ and from each other. Thus, by defining $y^* = f(x^*), y^{(j)} = f(x^{(j)})$ we see that these points conform a fracture of X in the direction $\psi(f)\boldsymbol{q}$.

The group $\operatorname{Sym}(X, \mathbb{Z}^2)$ is forced to act “naturally” over the set of directions of fracture; thus, constraints for these directions enforce similar restrictions on the possible values of $\psi(f)$ for $f \in \operatorname{Sym}(X, \mathbb{Z}^2)$. Note the analogy with bijective substitutions in the previous section.

Proposition 4.11 *For the Robinson shift,* $\operatorname{Sym}(X_{\text{Rob}}, \mathbb{Z}^2) \cong \mathbb{Z}^2 \rtimes D_4$*, where* $D_4 = Q_2$ *is the dihedral group of order* 8*, that is, the group of isometries of the square.*

Proof. To prove this result, we will show that the set $\mathcal{S}$ of all directions of fracture of X_{Rob} is $\mathbb{Z}\boldsymbol{e}_1 \cup \mathbb{Z}\boldsymbol{e}_2$. Assuming this as true, we see that, since $\psi(f)$ is always a $\mathbb{Z}$-invertible matrix, it must send $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ to a basis of $\mathbb{Z}^2$ contained in $\mathbb{Z}\boldsymbol{e}_1 \cup \mathbb{Z}\boldsymbol{e}_2$, which is always a two-element set of the form $\{\pm\boldsymbol{e}_1, \pm\boldsymbol{e}_2\}$ or $\{\pm\boldsymbol{e}_1, \mp\boldsymbol{e}_2\}$, and thus the elements of $\operatorname{Sym}(X_{\text{Rob}}, \mathbb{Z}^2)$ correspond to one of the eight possible matrices belonging to the standard copy of $D_4 = Q_2$ (defined in the previous section) in $\operatorname{GL}_2(\mathbb{Z})$. Then, by finding an explicit subgroup of $\operatorname{Sym}(X_{\text{Rob}}, \mathbb{Z}^2)$ isomorphic to D_4 by ψ, we deduce the claimed semidirect product decomposition.

To show that X_{Rob} has fractures in the directions $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$, we need to recall some basic details about the construction of an infinite valid configuration of the Robinson shift. As stated above, the five basic Robinson tiles (together with their rotations and reflections) combine to form 3×3 patterns with a similar behavior to crosses, named second order supertiles. Four of these second order supertiles, together with smaller substructures, further combine to form 7×7 patterns (third order supertiles) and so on. In every case, the central tile of an n-th order supertile is a cross, which gives an orientation to the supertile in a similar way to the two-headed, L-shaped arrow on a cross.

We may fill the whole upper right quadrant $Q_{\mathbb{1}} = \mathbb{N}^2$ as follows: we start by placing a cross on its vertex $\mathbf{0}$ with its L-shaped arrow pointing up and right, and then place another cross with the same orientation at the position $(1,1)$. This new cross, together with the previously placed one, allows us to fill the lower 3×3 section of $\mathbb{N}^2$, $[0,2]^2$, with a second order supertile. We iterate this process by placing a cross with the same orientation at the position $(3,3), (7,7), \dots, (2^n-1, 2^n-1), \dots$ and constructing the corresponding second, third, $\dots$ n-th order supertile and so on. By compactness, there is only one way to fill all of $\mathbb{N}^2$ as a limit to this process. We call the resulting configuration an infinite order supertile.

We may fill the other three quadrants with similar constructions resulting in infinite order supertiles with different orientations, each of these separated from the other infinite supertiles by a row or column of copies of the first tile from Figure 4.6. As we see in Figure 4.8, this will result in a translate of $(\mathbb{Z} \times \{0\}) \cup (\{0\} \times \mathbb{Z})$ containing only copies of this tile, with all of the tiles in one of the strips $\mathbb{Z} \times \{0\}$ or $\{0\} \times \mathbb{Z}$ (the latter in the figure) having the same orientation, while the other strip will have all of its tiles pointing towards the center.

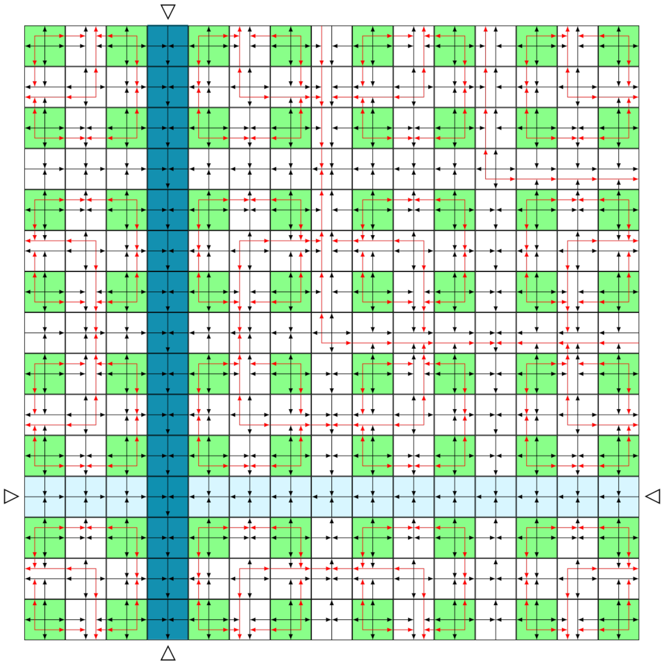


Figure 4.8: A fragment of a point from the Robinson shift, distinguishing the four supertiles involved, the vertical and horizontal strips of tiles separating each supertile and the $2\mathbb{Z} \times 2\mathbb{Z}$ sublattice that contains only crosses. Note that the tiles in the vertical strip separating the supertiles are copies of the first tile of Figure 4.6 with the same orientation.

Since the Robinson shift behaves like a nearest-neighbor shift with added restrictions, the existence of a vertical (resp., horizontal) strip with copies of the same tile allows us to vertically shift the tiles contained in the right half-plane however we see fit, as long as the coset of the $2\mathbb{Z} \times 2\mathbb{Z}$ sublattice containing only crosses is respected. In practice, this shows that in the point $x \in X_{\text{Rob}}$ represented partially in Figure 4.8 we may replace the tiles from the right half-plane with the corresponding tiles from $\sigma_{(0,2k)}(x)$ and obtain valid points[8] for all values of $k \in \mathbb{Z}$. We see an example of this in Figure 4.9.

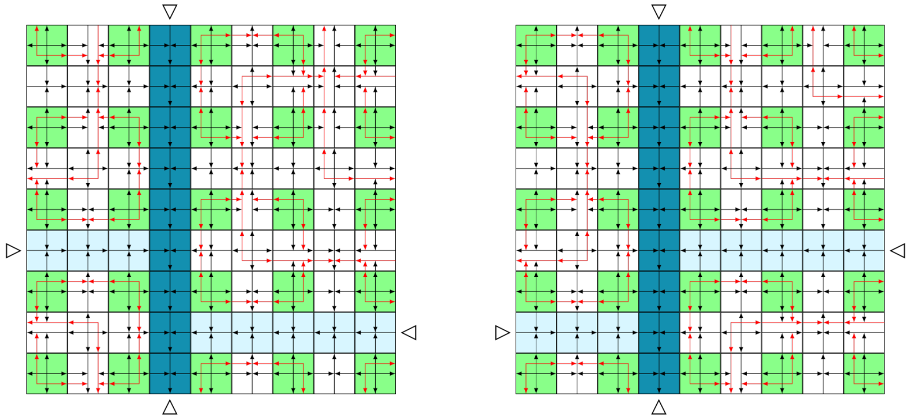


Figure 4.9: Two possible ways in which the tiling from Figure 4.8 exhibits fracture-like behavior, resulting in valid points from X_{Rob}.

This procedure shows that X_{Rob} has $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ as directions of fracture. Now, we need to show that all directions of fracture are contained in the set $\mathbb{Z}\boldsymbol{e}_1 \cup \mathbb{Z}\boldsymbol{e}_2$, and thus all matrices from $\psi[\text{Sym}(X_{\text{Rob}}, \mathbb{Z}^2)]$ send the set $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ to a linearly independent subset of $\{\boldsymbol{e}_1, \boldsymbol{e}_2, -\boldsymbol{e}_1, -\boldsymbol{e}_2\}$. The argument we shall use for this follows a similar outline to the technique used in the first half of the proof of Theorem 4.6: the points of the Robinson shift form a hierarchical structure away from a horizontal or vertical fracture, allowing for a decomposition into subpatterns of arbitrarily large size s placed correlative to a lattice of the form $2^n\mathbb{Z} \times 2^n\mathbb{Z}$ (this is similar to the decomposition of a point from a substitutive subshift into patterns of the form $\theta^m(a), a \in \mathcal{A}$ for arbitrarily large values of m). The existence of fractures that are neither vertical nor horizontal would result in "ruptures" in this hierarchical structure, leading to a contradiction.

Formally, we proceed as follows. Suppose that X_{Rob} has a fracture in the direction $\boldsymbol{q} \in \mathbb{Z}^2 \setminus (\mathbb{Z}\boldsymbol{e}_1 \cup \mathbb{Z}\boldsymbol{e}_2)$, and let S^+, S^- be the disjoint half-planes separated by $\boldsymbol{q}$. The set $F_{\boldsymbol{q}} = \mathbb{Z}^2 \setminus (S^+ \uplus S^-)$ is necessarily of the form $\mathbb{Z}\boldsymbol{q} + [\boldsymbol{r}_1, \boldsymbol{r}_2]$, namely, a finite union of translates of $\mathbb{Z}\boldsymbol{q}$, and thus its intersection with any set of the form $\mathbb{Z} \times \{k\}$ or $\{k\} \times \mathbb{Z}$ is finite. This is because the intersection of such a set with $\mathbb{Z}\boldsymbol{q}$ consists of at most a single point, as $\boldsymbol{q}$ is not a multiple of $\boldsymbol{e}_1$ nor $\boldsymbol{e}_2$. Thus, for any sufficiently large value $M \in \mathbb{N}$, it is easy to verify that for any point $\boldsymbol{p} \in F_{\boldsymbol{q}}$, any translation of the rectangle $[-M\mathbb{1}, M\mathbb{1}]$ that contains $\boldsymbol{p}$ also contains points from either S^+ or S^- (or both).

Choose $n \in \mathbb{N}, n > 1$, such that for $M = 2^n - 1$ the above condition holds, while satisfying the additional condition $M > 2k_1\|\boldsymbol{q}\|_1$. All n-th order supertiles thus contain points from

[8] Note that, while the "fractured" points do not appear in the minimal subset M_{Rob}, the originating four-supertile point does indeed belong to this minimal subshift, a detail which will be important in its study down below.

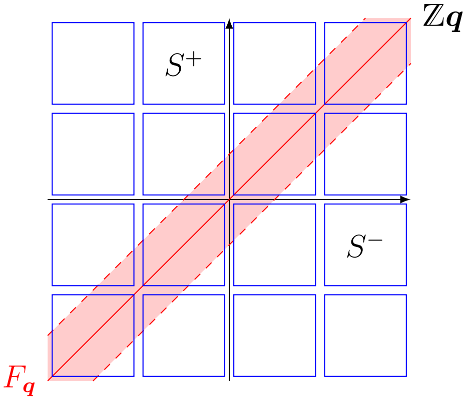


Figure 4.10: The substructure of a point of X_{Rob} in terms of n-th order supertiles. Note how all supertiles overlap either S^+ or S^-.

$S^+ \uplus S^-$. Let $\boldsymbol{p}$ be an element from $F_{\boldsymbol{q}}$ that belongs to the support of an n-th order supertile, and suppose this supertile overlaps the half-plane S^+. If there is no such supertile, all n-th order supertiles containing points of $F_{\boldsymbol{q}}$ only overlap S^-, implying, since S^+ is the intersection of a real half-plane $H_{\boldsymbol{\alpha},c} = \{\boldsymbol{v} \in \mathbb{R}^2 : \langle \boldsymbol{v}, \boldsymbol{\alpha}\rangle \geq c\}$ with $\mathbb{Z}^2$, that S^+ is a translation of $\mathbb{Z}\times(\pm\mathbb{N})$ (or $(\pm\mathbb{N})\times\mathbb{Z}$). Convex combinations of points of S^+ with integer coefficients belong to S^+ as well, so S^+ cannot have "gaps", and it is a union of disjoint, horizontally or vertically adjacent translates of $[1, 2^n]^2$. This implies that $\boldsymbol{q}$ is in the set $\mathbb{Z}\boldsymbol{e}_1 \cup \mathbb{Z}\boldsymbol{e}_2$, a contradiction. Thus, the aforementioned supertile exists. Evidently, the same argument shows the existence of other n-th order supertiles which intersect S^-.

Since each horizontal or vertical strip $F_{\boldsymbol{q}} \cap (\mathbb{Z}\times\{k\})$ (resp. $F_{\boldsymbol{q}} \cap (\mathbb{Z}\times\{k\})$) intersects finitely many supertiles, we see that the arrangement of the n-th order supertiles in S^+ away from a vertical or horizontal fracture (which in this case must correspond to a bi-infinite column or row of copies of the first tile from Figure 4.6, all with the same orientation) affects the placement of the supertiles in S^- as well. However, since the tiling has a fracture in the direction $\boldsymbol{q}$, we may shift the supertiles in S^- by $k_1\boldsymbol{q}$ and obtain a valid configuration. By our choice of n, the shift $\sigma_{k_1\boldsymbol{q}}$ moves the n-th order supertiles by less than M units both horizontally and vertically (since $M > 2k_1\|\boldsymbol{q}\|_1$), and thus the supertiles in S^- are shifted to a position that does not match the arrangement of supertiles from S^+. We may see this situation in Figure 4.11.

Given that we are assuming that this point (say, x) is a fracture point for X_{Rob}, there must be some other point y which matches x in S^+ and $\sigma_{k_1\boldsymbol{q}}(x)$ in S^-, which breaks the rigidity of the structure of supertiles imposed by the rules of the Robinson shift. Thus, fractures along non-principal directions cannot exist.

Finally, we need to construct a copy of D_4 contained in $\operatorname{Sym}(X_{\text{Rob}}, \mathbb{Z}^2)$. For this, since D_4 is a 2-generated group, we only need to show the existence of two extended symmetries $\rho, \mu : X_{\text{Rob}} \to X_{\text{Rob}}$, mapped respectively by ψ to the matrices:

$$\psi(\rho) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \psi(\mu) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix},$$

since these two matrices generate an isomorphic copy of D_4 contained in $\mathrm{GL}_2(\mathbb{Z})$. These symmetries ρ and μ are essentially rigid symmetries of the coordinate axes; however, a com-

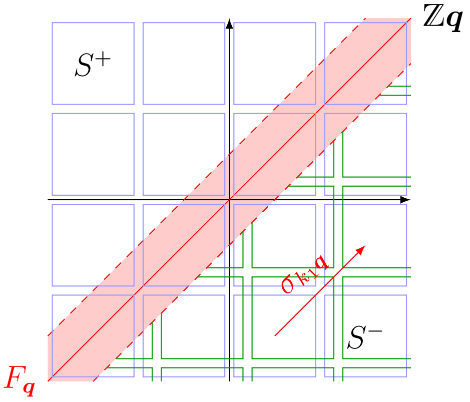


Figure 4.11: How a shift by $k_1\boldsymbol{q}$ makes the arrangement of supertiles in S^+ not match with the corresponding tiles in S^-.

position with a relabeling map is also needed, to replace every tile with the corresponding reflection or rotation. For instance, if we define $\mathfrak{R} : \mathcal{A} \to \mathcal{A}$ as the mapping which assigns to each of the 28 symbols its corresponding rotation by $\frac{1}{2}\pi$, as seen in Figure 4.12, then $\rho(x)_{(i,j)} = \mathfrak{R}(x_{(-j,i)})$ is the desired symmetry. In the same way, by defining $\mathfrak{M} : \mathcal{A} \to \mathcal{A}$ as the mapping that sends each tile to its reflection through the horizontal axes, then we define μ by the relation $\mu(x)_{(i,j)} = \mathfrak{M}(x_{(-i,j)})$.

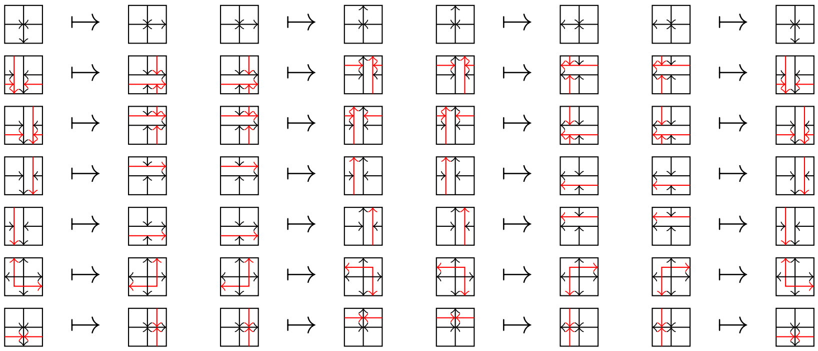


Figure 4.12: The relabeling map $\mathfrak{R}$ which replaces each tile with its corresponding rotation by $\frac{1}{2}\pi$.

It is easy to verify that ρ and μ are valid extended symmetries, as they respect the conditions on the arrowheads and tails, and the sublattice comprised of only crosses. Also, we see that ψ sends both ρ and μ to the desired matrices, and that the mappings $\mathfrak{R}_\infty, \mathfrak{M}_\infty : \mathcal{A}^{\mathbb{Z}^2} \to \mathcal{A}^{\mathbb{Z}^2}$ commute with the corresponding rigid symmetries of the coordinate axes. Thus, $\langle \rho, \mu \rangle$ is a copy of D_4 contained in $\operatorname{Sym}(X_{\mathrm{Rob}}, \mathbb{Z}^2)$, as desired.

We remark that the proof above used the structure of the Robinson shift X_{Rob} exclusively to compute the set of directions of fractures associated to this shift, and that extended symmetries preserve this set in other contexts as well. This suggests that this technique is open to generalization to other subshifts, even in higher dimensions, although possibly replacing the concept of "direction of fracture" with "hyperplane of fracture", as we need to separate half-spaces of $\mathbb{Z}^d$. We suggest the following tentative definition for a fracture in a d-dimensional subshift:

Definition 4.12 *Let X be a (strongly aperiodic) $\mathbb{Z}^d$-subshift. We say that X has a* ***fracture*** *in the direction of the hyperplane $H = H_0 + \boldsymbol{v}$ if for some $x \in X$ there are two half-spaces S^+, S^- separated by H (i.e. $S^+ \cap S^- = S^+ \cap H = S^- \cap H = \varnothing$) such that for some "sufficiently large" subset $B \subseteq H_0$ there is a family $\{x^{(\boldsymbol{b})}\}_{\boldsymbol{b} \in B}$ of points of X such that:*

$$x^{(\boldsymbol{b})}|_{S^+} = x|_{S^+}, \qquad x^{(\boldsymbol{b})}|_{S^-} = \sigma_{\boldsymbol{b}}(x)|_{S^-}.$$

Here, an appropiate definition of "sufficiently large" will depend on the subshift that is being studied. For instance, in the case of the Robinson shift we only needed B to contain two points ($\{0, k_1\boldsymbol{q}\}$) for our argument due to the hierarchical structure of X_{Rob}, albeit B in this shift actually is an infinite set, $2\mathbb{Z}\boldsymbol{q}$. In all cases, as long as we apply a consistent restriction to the possible instances of B, we see that an extended symmetry f must send a point of fracture to another point of fracture due to the generalized CHL theorem, and thus $\psi(f)$ is a matrix that acts by permutation on the set of all hyperplanes of fracture of X. For a sufficiently rigid $\mathbb{Z}^d$-shift X, this should result in a strong restriction on the matrix group $\psi[\text{Sym}(X, \mathbb{Z}^d)]$.

It is important to note the analogy between the method used for the Robinson tiling and the ideas discussed before for bijective substitutions, and the above generalization scheme for higher-dimensional fractures serves to highlight this similarity. We expect that many symbolic systems with a strong hierarchical structure where substructures of large support can be built "independently" should be amiable to such methods, as in the above two situations.

The discussion above shows the key idea behind the method: the hierarchical structure of the aforementioned subshifts forces the appearance of "special directions", which result in a geometrical invariant that needs to be preserved by extended symmetries. By identifying these special directions via combinatorial or dynamic properties, we can effectively restrict $\psi[\text{Sym}(X, \mathbb{Z}^d)]$ enough to effectively compute it in terms of $\text{Aut}(X, \mathbb{Z}^d)$.

However, as stated before, the Robinson shift X_{Rob} is not minimal. To exhibit the above mentioned special directions, a key point was using certain points that exhibit "fracture-like" behavior, which are not present in the minimal subset M_{Rob}. However, since the "special directions" come from the hierarchical structure of the subshift, they ought to be present in its minimal subset as well in some form, and thus, they should impose the same restrictions on the set of extended symmetries. We conclude this discussion by showcasing a method to exhibit these directions in the minimal subset of the Robinson shift, thus proving that it has the same extended symmetry group as its standard counterpart. The argument is as follows:

Corollary 4.13 *Let $M_{\text{Rob}} \subset X_{\text{Rob}}$ be the unique minimal subshift contained in X_{Rob}. Then, $\text{Sym}(M_{\text{Rob}}, \mathbb{Z}^2) \cong \mathbb{Z}^2 \rtimes D_4$.*

Proof. Using the substitution rules devised by Gähler in [43], we can show that M_{Rob} contains a point, say x, that has only copies of the first tile from Figure 4.6, pointing to the right, on the horizontal strip $\mathbb{Z} \times \{0\}$, and corresponding tiles of the same kind pointing downwards in $\{0\} \times \mathbb{Z}^+$ and upwards in $\{0\} \times \mathbb{Z}^-$. Mirrored and rotated versions of this configuration exist as points of M_{Rob} as well (of which one specific rotation may be observed in Figure 4.8); a similar argument holds for the fifth tile.

Any point from $\mathcal{H} = \overline{\{\sigma_{(n,0)}(x) : n \in \mathbb{Z}\}}$ has the same horizontal strip of copies of the same tile on $\mathbb{Z} \times \{0\}$, and, due to the local rules of the Robinson tiling, any configuration with support $\mathbb{Z} \times [-n, n]$ from some point $y \in \mathcal{H}$ must be $(m, 0)$-periodic for some sufficiently large m. Note that this m must diverge to ∞ as $n \to \infty$, because no point from M_{Rob} has nontrivial periods.

Let $f \in \text{Sym}(M_{\text{Rob}}, \mathbb{Z}^2)$ be an extended symmetry. For any sufficiently large value of $k \in \mathbb{N}$, the window of this f is contained in $\mathbb{Z} \times [-k, k]$. Thus, due to Theorem 3.10, we may choose a sufficiently large k such that the image of $\mathbb{Z} \times [-k, k]$ under the matrix $\psi(f)$ contains the set $L_{a,b}(\widetilde{k}) := \{(u, v) \in \mathbb{Z}^2 : -\tilde{k} \leq au + bv \leq \tilde{k}\}$ for any desired $\tilde{k} > 0$ and some $a, b \in \mathbb{Z}$, and thus $y|_{\mathbb{Z}\times[-n,n]}$ determines $f(y)|_{L_{a,b}(\widetilde{k})}$ entirely. Since the strip $y|_{\mathbb{Z}\times[-k,k]}$ is periodic, the restriction $f(y)|_{L_{a,b}(\tilde{k})}$ must have a period as well, which we can choose as a multiple of $(-b, a)$.

Suppose that $ba \neq 0$, which implies that $\psi(f)$ maps $\boldsymbol{e}_1$ to a direction that is not parallel to the coordinate axes. Since the n-th order supertiles increase in size exponentially with n, and so do the associated "square drawings" determined by the crosses, the strip $L_{a,b}(\widetilde{k})$ must pass through the vertical lines (comprised of copies of rotations of the second, third, fourth or fifth tiles from Figure 4.6) associated with the corresponding square of a n-th order supertile for all sufficiently large n (as it is not parallel to any of the sides of such squares). Thus, this configuration cannot have a nontrivial period, since due to the positions of the n-th order supertiles such a period cannot have a horizontal or vertical component smaller than 2^n, which applies for any sufficiently large n. We conclude, by this contradiction, that $\psi(f)$ maps $\boldsymbol{e}_1$ to a vector parallel to the coordinate axes; a similar argument holds with $\boldsymbol{e}_2$.

Remark We may also proceed with a topological method similar to the one mentioned in Remark 4.3. As shown in [43], M_{Rob} factors onto a two-dimensional **solenoid**[9] $\mathbb{S}_2^2$, and this factor map is 28-to-1 in the set of all points from M_{Rob} comprised of four infinite-order supertiles, which is thus preseved by any extended symmetry, as can be seen from a fiber cardinality argument. We arrive to the same conclusions after appealing to the CHL theorem.

Acknowledgements

The author would like to thank his thesis advisor, Michael H. Schraudner, for his guidance, advice, insight and ideas with regard to the development of the current work, and Michael Baake and Franz Gähler from Bielefeld University for their helpful comments and advice regarding further developments and additions. As well, the author thanks the referees who revised earlier versions of this paper and offered many insightful suggestions to improve its presentation.

The author would also like to kindly acknowledge the financial support received from ANID, via Doctoral Fellowship ANID-PFCHA/Doctorado Nacional/2017-21171061. Significant revisions to the original form of this paper were enacted during a stay at Bielefeld, which was

[9] The solenoid $\mathbb{S}_p$ is the compact abelian group obtained as an inverse limit of the system $\mathbb{R}/\mathbb{Z} \leftarrow \mathbb{R}/\mathbb{Z} \leftarrow \mathbb{R}/\mathbb{Z} \leftarrow \dots$, where each group homomorphism is the mapping $x \mapsto px \pmod 1$. A d-dimensional solenoid is defined analogously.

financed by the CRC 1283 at Bielefeld University. Additional financial support for this stay was received from University of Chile, under the project UCH-1566.

Chapter 5

Admissible extended symmetries and geometrical constraints

This chapter is an adaptation of the publication *Admissible Reversing and Extended Symmetries for Bijective Substitutions* [23], which is a joint work with Daniel Luz[1] and Neil Mañibo[2] from Universität Bielefeld.

5.1. Introduction

The study of automorphism groups, often also known as symmetry groups, is an important part of the analysis of a dynamical system, as it can offer insight on the behaviour of the system, as well as allowing classifications of distinct families of dynamical systems (acting as a conjugacy invariant). In particular, automorphism groups of shift spaces have been thoroughly studied (see e.g. the analysis of the automorphism group of the full shift [19], the series of works on automorphisms in low-complexity subshifts [29, 31, 33], and recent works on shifts of algebraic and number-theoretic origin [4, 39]).

Automorphisms of subshifts can be algebraically defined as elements of the topological centralizer of the group $\langle\sigma\rangle$ generated by the shift, seen as a subgroup of the space $\mathrm{Aut}(X)$ of all self-homeomorphisms of X onto itself. Thus, a natural question at this point is whether the corresponding normalizer has an interesting dynamical interpretation as well. This leads to the concept of **reversing symmetries** (for $d = 1$); see [12, 13, 47], the monograph [82] for a group-theoretic exposition, and [67] for a more physical background. These are special types of flip conjugacies; see [15]. In higher dimensions, one talks of **extended symmetries**; see [3, 13], which are examples of $\mathrm{GL}(d, \mathbb{Z})$-conjugacies; compare [4, 66]. These kinds of maps are related to phenomena such as palindromicity and several properties of geometric and topological nature, which is more evident in the higher-dimensional setting [13, 22].

High complexity is often (but not always, see for instance the square-free subshift [4]) linked to a complicated automorphism group. For instance, determining whether the automorphism groups of the full shifts in two and three symbols are isomorphic has consistently proven to be a difficult question [19]. The low-complexity situation, thus, often allows for a more

[1] dluz@math.uni-bielefeld.de

[2] cmanibo@math.uni-bielefeld.de

in-depth analysis and more complete descriptions, up to and including explicit computation of these groups in many cases.

The particular case of substitutive subshifts has gathered significant attention and here a lot of progress has been made; see [57, 80]. Unsurprisingly, the presence of non-trivial automorphisms is also tied to the spectral structure of the underlying dynamical system; see [41, 85]. In this work, we restrict to systems generated by bijective substitutions, both in one and in higher dimensions. These substitutions are typically n-to-1 extensions of odometers and generate coloured tilings of $\mathbb{Z}^d$ by unit cubes, where one usually identifies a letter with a unique colour; see [41]. We compile and extend known properties about this family of substitutive subshifts regarding automorphisms. Some natural questions in this direction are:

1. What kinds of groups can appear as automorphism or extended symmetry groups of specific substitutive subshifts?
2. Given a specific group G, can we construct a substitution whose associated subshift has G as its automorphism or extended symmetry group?

Both questions are accessible for bijective substitutions. For automorphism groups, the second question is answered in full in [33], which extends to higher dimensions with no additional assumptions because the result does not depend on the geometry of the substitution; see [27] for realisation results for more general group actions. We add to such known results in Theorem 5.9. Aperiodicity also plays a key role here, which can easily be confirmed in the bijective setting; see Propositions 5.4 and 5.21.

On the other hand, the existence of non-trivial reversing or extended symmetries depends heavily on the geometry and requires more in terms of the relative positions of the permutations in the corresponding supertiles, the expansive maps, and the shape of the supertiles themselves. In Theorem 5.13, we provide equivalent conditions for the existence of non-trivial reversing symmetries, which we generalise to higher dimensions in Theorem 5.19 to cover extended symmetries.

As a corollary, in any dimension d, given a finite group G and a subgroup P of the hyperoctahedral group P, we provide a construction in Theorem 5.22 of a bijective substitution whose underlying shift space has automorphism and extended symmetry group $\mathbb{Z}^d \times G$ and $(\mathbb{Z}^d \rtimes P) \times G$, respectively. A similar construction with a different structure of the extended symmetry group is done in Theorem 5.23. We also provide algorithms on how one can check whether there exist non-trivial automorphisms and extended symmetries for a given substitution θ; see Sections 5.2.2 and 5.3.1.

5.2. Bijective constant-length substitutions

5.2.1. Setting and basic properties

Let $\mathcal{A}$ be a finite alphabet and $\mathcal{A}^+ = \bigcup_{L\geq 1} \mathcal{A}^L$ be the set of finite non-empty words over $\mathcal{A}$; we shall write $\mathcal{A}^* = \mathcal{A}^+ \cup \{\varepsilon\}$, where the latter is the empty word. As in previous chapters, a (one-dimensional) **substitution** is a map $\theta\colon \mathcal{A} \to \mathcal{A}^+$; **constant-length substitutions** are those for which, for some $L \in \mathbb{N}$, all words have the same length L. As in the previous chapter, we focus on primitive substitutions.

Given a primitive substitution θ, we once again mention that the set of all words (patterns) that appear as subwords of some word (respectively, pattern) of the form $\theta^k(a)$ for some $a \in \mathcal{A}, k \geq 1$ is extensible and factorial, and thus entirely defines a shift space, which is the substitutive subshift X_θ defined previously [41, 65, 85]. It is well known that the primitivity of θ implies that X_θ is strictly ergodic (uniquely ergodic and minimal); see [6, 85]. We refer the reader to [75] for a treatment of substitutions which are non-primitive.

As defined before, a constant-length substitution $\theta\colon \mathcal{A} \to \mathcal{A}^L$ is **bijective** if the map $\theta_j\colon a \mapsto \theta(a)_j$ is a bijection on $\mathcal{A}$, for all indices $0 \leq j \leq L-1$. Equivalently, θ is bijective if there exist L (not necessarily distinct) bijections $\theta_0, \ldots, \theta_{L-1}\colon \mathcal{A} \to \mathcal{A}$ such that $\theta(a) = \theta_0(a) \ldots \theta_{L-1}(a)$ for every $a \in \mathcal{A}$. We shall refer to the mapping θ_j as the j-th **column** of the substitution θ.

Consider $\{\theta_j\}_{j=0}^{L-1} \subset S_{|\mathcal{A}|}$. Let $\Phi\colon S_{|\mathcal{A}|} \to \mathrm{GL}(|\mathcal{A}|, \mathbb{Z})$ be the representation via permutation matrices. One then has the following; compare [41, Cor. 1.2].

Fact 5.1 *Let θ be a primitive, bijective substitution, with columns $\{\theta_0, \ldots, \theta_{L-1}\}$. Then the substitution matrix M is given by $M = \sum_{j=0}^{L-1} \Phi(\theta_j^{-1})$. Moreover, $(1, 1, \ldots, 1)^T$ is a right Perron–Frobenius eigenvector of M, so each letter has the same frequency for every element in the hull X_θ, i.e., $\nu_a = \frac{1}{|\mathcal{A}|}$ for all $a \in \mathcal{A}$ and all $x \in \mathsf{X}_\theta$.* □

Define the n-th **column group** $G^{(n)}$ to be the following subgroup of the symmetric group of bijections $\mathcal{A} \to \mathcal{A}$:

$$G^{(n)} := \langle \{\theta_{j_1} \circ \cdots \circ \theta_{j_n} \ : \ 0 \leq j_1, \ldots, j_n \leq L-1\} \rangle.$$

As it turns out, the groups $G^{(n)}$ generated by the columns give a good description of the substitution θ in the bijective case; see [57] for its relation to the corresponding Ellis semigroup of X_θ. The primitivity of θ may be characterised entirely by this family of groups, as seen below. Recall that a subgroup $G \leq S_n$ of the symmetric group on $\{1, \ldots, n\}$ is **transitive** if for all $1 \leq j, k \leq n$ there exists $\tau \in G$ such that $\tau(j) = k$. Here, we let $N \in \mathbb{N}$ be the minimal power such that $\theta_j^N = \mathrm{id}$ for some $0 \leq j \leq L^N - 1$; compare [85, Lem. 8.1]. In [57], $G^{(N)}$ is called the **structure group** of θ.

Proposition 5.2 *Let $\theta\colon \mathcal{A} \to \mathcal{A}^L$ be a bijective substitution. Then, the following are equivalent:*

1. *The substitution θ is primitive.*
2. *All groups $G^{(n)}, n \in \mathbb{N}$, are transitive.*
3. *The group $G^{(N)}$ is transitive.*

Proof. Evidently, (2) $\implies$ (3), so we only need to prove (3) $\implies$ (1) $\implies$ (2). To see the first implication, note first that the columns of the iterated substitution θ^N are compositions of the form $\theta_{j_1,\ldots,j_N} := \theta_{j_1} \circ \cdots \circ \theta_{j_N}, 0 \leq j_1, \ldots, j_N \leq L-1$, that is, for any $a \in \mathcal{A}$ the following holds:

$$\theta^N(a) = \theta_{0,\ldots,0,0}(a)\theta_{0,\ldots,0,1}(a) \ldots \theta_{0,\ldots,0,L-1}(a)\theta_{0,\ldots,1,0}(a) \ldots \theta_{L-1,\ldots,L-1,L-1}(a).$$

Since, by (3), the group $G^{(N)}$ is transitive, the substitution matrix M_{θ^N} is irreducible, i.e. it is the adjacency matrix of a strongly connected digraph. In other words, for all $a, b \in \mathcal{A}$, there exists a composition of columns $q, q', \dots, q''$ of θ^N such that $q \circ q' \circ \cdots \circ q''(a) = b$, which may be identified with a path in the graph whose vertices are the letters of $\mathcal{A}$ and with one edge from c to $r(c)$ for any $c \in \mathcal{A}$ and column r. The choice of N also shows that M_{θ^N} has a non-zero diagonal, since one of the columns of θ^N is the identity. These two conditions immediately imply that M_{θ^N} is a primitive matrix (see [71, Ch. 2]) which in turn implies primitivity of θ, as desired.

To prove (1) $\implies$ (2), note that primitivity of θ implies that, for some $k > 0$ and for all $a \in \mathcal{A}$, the word $\theta^k(a)$ contains all symbols of the alphabet $\mathcal{A}$, including a itself. Since the columns of θ^k generate $G^{(k)}$, this implies that for all $a, b \in \mathcal{A}$ there is some generator of this group that maps a to b, i.e. $G^{(k)}$ is transitive. Since $\theta^k(a)$ contains a as a subword, this implies that $\theta^{2k}(a)$ contains $\theta^k(a)$ as a subword, and, by induction, that $\theta^{mk}(a)$ contains $\theta^k(a)$ as a subword for all $m \geq 1$; thus, all groups $G^{(mk)}$ are transitive. Now, it is easy to see that $G^{(n)} \leq G^{(d)}$ if $d \mid n$. Then, for all $n \in \mathbb{N}$, $G^{(n)}$ has $G^{(nk)}$ as a transitive subgroup and hence it is transitive.

The bijective structure of θ can also be exploited to conclude the aperiodicity of X_θ by just looking at simple features of θ. Below, we provide several criteria for aperiodicity in terms of $|\mathcal{A}|$, L, and the existence of certain legal words.

Proposition 5.3 *Let X_θ be the subshift generated by a primitive, bijective substitution θ of length L on a finite alphabet $\mathcal{A}$. If* $\gcd(|\mathcal{A}|, L) > 1$ *then X_θ is aperiodic.*

Proof. Assume that w^∞ is a periodic word in X_θ with least period p, i.e., $w^\infty = v^\infty$ with v being a prime period ($|v| = p$). Then without loss of generality, we assume that w^∞ is fixed under θ by replacing it with a power θ^k such that the first column of θ^k is the identity. We choose the smallest possible constants $c, d \in \mathbb{N}$ which satisfy $cL = dp$. That is, the word $w^\infty|_{[0,cL-1]}$ is an inflation of c letters and, at the same time, d copies of the prime period. Since θ is a bijective substitution of length L, every inflation word of length cL has exactly one preimage under θ, which is a word of length c. In particular, since w^∞ is fixed under θ, the preimage of $w^\infty|_{[0,cL-1]}$ under θ must be an initial segment $x_1 \dots x_c$ of w^∞ of length c. As cL is a multiple of p, then, for any $k \in \mathbb{Z}$, $w^\infty|_{[0,cL-1]} = \left(\sigma^{kcL}(w^\infty)\right)_{[0,cL-1]}$, which all have the same preimage under θ. This means that w^∞ is an infinite concatenation of copies of $x_1 \dots x_c$ and is thus c-periodic. As p is the least period, we must have $c = ep$ for some integer e. Since c is minimal $c = p$ and thus $d = L$, which certainly solves $cL = dp$.

From Fact 5.1 we know that every letter has the same frequency for any element in X_θ. This, together with the fact that w^∞ is a concatenation of v, implies that every letter appears equally often within v, so $|\mathcal{A}| \mid p$. If $\gcd(|\mathcal{A}|, L) > 1$ then $\gcd(p, L) = a > 1$ as well. But then $\frac{c}{a}L = \frac{d}{a}p$ holds and $c' = \frac{c}{a}$ and $d' = \frac{d}{a}$ are smaller integer constants contradicting the minimality of c and d. So our assumption that w^∞ is periodic has to be false.

Another way to get aperiodicity is through the existence of proximal pairs; see [33, Sec. 3.2.1] and [6, Cor. 4.2 and Thm. 5.1]. Two elements $x \neq y \in (X, \sigma)$ are said to be proximal if there exists a subsequence $\{n_k\}$ of $\mathbb{N}$ or $-\mathbb{N}$ such that $\mathrm{d}(\sigma^{n_k}x, \sigma^{n_k}y) \to 0$ as $k \to \infty$. A stronger notion is that of asymptoticity, which requires $\mathrm{d}(\sigma^n x, \sigma^n y) \to 0$ as $n \to \infty$ or $-\infty$. For bijective substitutions, these two notions are equivalent, and asymptotic pairs are completely characterised by fixed points of θ; see [57].

Consider a one-dimensional substitution θ and a fixed point w arising from a legal **seed**

$a|b$, i.e., $w = \theta^{\infty}(a|b)$. Here, the vertical bar represents the location of the origin, and the letter a generates all the letters at the negative positions, while b does the same for all non-negative ones. Two fixed points $w_1, w_2 \in \mathsf{X}_\theta$ generated by $a_1|b_1$ and $a_2|b_2$ are **left-asymptotic** if they agree at all negative positions and disagree for all non-negative positions. Right-asymptotic pairs are defined in a similar manner. We have the following equivalent condition for aperiodicity in terms of existence of certain legal words; compare [57, Prop. 4.1]

Proposition 5.4 *Let θ be a primitive, bijective substitution on a finite alphabet $\mathcal{A}$ in one dimension. Then the hull X_θ is aperiodic if and only if there exist distinct legal words of length 2 which either share the same starting or ending letter.*

Proof. Let $\theta := \theta_0 \cdots \theta_{L-1}$, with $\theta_i \in G$. Choosing $k = \operatorname{lcm}(|\theta_0|, |\theta_{L-1}|)$, we get that the first and the last columns of θ^k are both the identity, i.e., $\theta_0^k(a) = \theta_{L^k-1}^k(a) = a$ for all $a \in \mathcal{A}$. If there exist $ab, ac \in \mathcal{L}_\theta$ with $b \neq c$, the bi-infinite fixed points $\theta^{\infty}(a|b)$ and $\theta^{\infty}(a|c)$ they generate under θ^k coincide in all negative positions and differ in at least one non-negative position, and hence are left-asymptotic and proximal. Since X_θ is minimal and admits a proximal pair, all of its elements must then be aperiodic. Now suppose that every letter has a unique predecessor and successor in $\mathcal{A}$. This means that every element $x \in \mathsf{X}_\theta$ is uniquely determined by the letter at the origin. From the finiteness of $\mathcal{A}$, one gets $x = w^{\infty}$ and hence is periodic, from which the periodicity of the hull follows.

Example The substitution $\theta\colon a \mapsto aba, b \mapsto bab$ is primitive, bijective and admits a periodic hull. Here, the only legal words of length 2 are ab and ba. Note that θ is of height 2 and generates the same hull as the substitution $\theta'\colon a, b \mapsto ab$.

5.2.2. Automorphisms

In the following sections, we deal with the **automorphism group**[3] of our subshifts of interest, which are certain homeomorphisms of the shift space which preserve the dynamics of the shift action in a specific sense. Remember that this group is the set of all homeomorphisms $f\colon X \to X$ which commute with the shift action:

$$(\forall \boldsymbol{n} \in \mathbb{Z}^d)\colon \sigma_{\boldsymbol{n}} \circ f = f \circ \sigma_{\boldsymbol{n}}. \tag{5.1}$$

That is, $\operatorname{Aut}(X, \mathbb{Z}^d)$ is the centralizer of the set of shift maps in the group of all self-homeomorphisms of the space X. As already stated in previous chapters, f is totally determined by its **local function** $F\colon \mathcal{A}^U \to \mathcal{A}$, with $U \Subset \mathbb{Z}^d$ finite (see Theorem 3.10).

Automorphism groups of one-dimensional bijective substitutions are a thoroughly studied subject, both in the topological and ergodic-theoretical contexts. Complete characterisations of these groups are known, as seen in e.g. [28] for a two-symbol alphabet, or [70] for a characterisation in the measurable case; see also [29, 41] for further elaboration in the description of the automorphisms in this category of subshifts. The following theorem summarizes this classification:

[3] This group is called **symmetry group** in several sources, particularly those that follow the Smale convention, in which the notation $\operatorname{Aut}(X)$ refers to the set of all homeomorphisms, irregardless of whether they preserve the shift action. This includes the paper by Baake, Roberts and Yassawi [13] and the original, submitted version of this work.

Theorem 5.5 *Let X_θ be the hull generated by an aperiodic, primitive, bijective substitution θ on $\mathbb{Z}^d$. Then, the automorphism group* $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ *is isomorphic to the direct product of $\mathbb{Z}^d$, generated by the shift action, with a finite group of radius-0 sliding block codes $\tau_\infty \colon \mathsf{X}_\theta \to \mathsf{X}_\theta$ given by $\tau_\infty((x_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathbb{Z}^d}) = (\tau(x_{\boldsymbol{j}}))_{\boldsymbol{j}\in\mathbb{Z}^d}$ for some bijection $\tau \colon \mathcal{A} \to \mathcal{A}$.*

Furthermore, let N be any integer such that $\theta_{\boldsymbol{j}}^N$ is the identity for some $\boldsymbol{j}$ (note that such an N always exists). Then, $\tau \colon \mathcal{A} \to \mathcal{A}$ induces an automorphism if and only if $\tau \in \mathrm{cent}_{S_{|\mathcal{A}|}} G^{(N)}$. □

As a consequence, every automorphism on X_θ is a composition of a shift map and a radius-zero sliding block code as above. These conditions arise as a consequence of such a automorphism having to preserve the supertile structure of any $x \in \mathsf{X}_\theta$ at every scale, which in particular implies that a level-k supertile $\theta^k(a), a \in \mathcal{A}$ has to be mapped to some $\theta^k(b)$ for some other $b \in \mathcal{A}$ by the "letter exchange map" τ. The choice of N above ensures that, when k is a multiple of N, the equality $a = b$ holds, which implies that τ commutes with the columns of θ^N, and thus $\theta^N \circ \tau_\infty = \tau_\infty \circ \theta^N$. This in turn implies Eq. (5.1). For further elaboration on the proof of the above result, the reader may consult [29, 41], among others.

Example Consider the following substitution θ on the three-letter alphabet $\mathcal{A} = \{a, b, c\}$:

$$\begin{aligned} \theta \colon a &\mapsto abc, \\ b &\mapsto bca, \\ c &\mapsto cab. \end{aligned}$$

The columns correspond to the three elements of the cyclic group generated by $\tau = (a\,b\,c)$. It is not hard to verify that the only elements of $S_3 = D_3$ that commute with τ are the powers of τ themselves, and thus $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}) \cong \mathbb{Z} \times C_3$, with the finite subgroup C_3 being generated by the automorphisms induced by the powers of τ.

As it turns out, Theorem 5.5 provides an algorithm to compute $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z})$ explicitly. To introduce this algorithm, let us recall some easily verifiable facts from group theory [49, Ch. 1 and 5]:

Fact 5.6 *Let G be any group and $H = \langle S \rangle \leq G$ a subgroup generated by $S \subset G$. Then,*

$$\mathrm{cent}_G(H) = \{c \in G \mid (\forall h \in H) \colon ch = hc\} = \bigcap_{s \in S} \mathrm{cent}_G(s).$$ □

Fact 5.7 *Any permutation decomposes uniquely (up to reordering) as a product of disjoint cycles. Conjugation by some $\tau \in S_n$ can be computed from this decomposition using the identity:*

$$\tau(a_1\, a_2\, \ldots\, a_n)\tau^{-1} = (\tau(a_1)\, \tau(a_2)\, \ldots\, \tau(a_n)).$$

A permutation $\tau \in S_n$ belongs to $\mathrm{cent}_{S_n}(\pi)$ if and only if $\tau\pi\tau^{-1} = \pi$, and thus:

$$\begin{aligned} \pi &= (a_1\, a_2\, \ldots\, a_{k_1})(b_1\, b_2\, \ldots\, b_{k_2}) \cdots (c_1\, c_2\, \ldots\, c_{k_r}) \\ &= (\tau(a_1)\, \tau(a_2)\, \ldots\, \tau(a_{k_1}))(\tau(b_1)\, \tau(b_2)\, \ldots\, \tau(b_{k_2})) \cdots (\tau(c_1)\, \tau(c_2)\, \ldots\, \tau(c_{k_r})). \end{aligned}$$

Hence, the uniqueness of this decomposition implies that every cycle in the second decomposition is equal to a cycle of the same length in the first one. □

Thus, to compute the letter exchange maps that determine $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z})$, we need to find all permutations τ that preserve certain cycle decompositions. We obtain the following procedure:

Algorithm. Assuming that θ is a primitive, bijective, aperiodic substitution, the following algorithm computes $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z})$ explicitly.

- **Input:** θ is a length-L bijective substitution, which may be represented as a function (dictionary) $\theta \colon \mathcal{A} \to \mathcal{A}^L$ or a set of L permutations $\theta_0, \theta_1, \dots, \theta_{L-1} \colon \mathcal{A} \to \mathcal{A}$, corresponding to each column.
- **Output:** A (finite) set of permutations C forming a group, so that $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d) = \mathbb{Z}^d \times C$.

(1) Compute the least positive integer N such that $\theta_{\boldsymbol{j}}^N$ is the identity on $\mathcal{A}$ for some column of the substitution θ. N equals the least common multiple of all cycle lengths in the decomposition of the columns $\theta_{\boldsymbol{j}}$ into disjoint cycles (and is thus finite).

(2) Determine all columns $\theta_{\boldsymbol{j}_1} \circ \cdots \circ \theta_{\boldsymbol{j}_N}$ of the iterated substitution θ^N. This is a generating set for the group $G^{(N)}$.

(3) For every column computed in (2), compute $G_{\boldsymbol{j}_1,\dots,\boldsymbol{j}_N} = \mathrm{cent}_{S_n}(\theta_{\boldsymbol{j}_1} \circ \cdots \circ \theta_{\boldsymbol{j}_N})$ by taking the cycle decomposition of this permutation (in where we identify $\mathcal{A}$ with the set $\{1, 2, \dots, |\mathcal{A}|\}$) and employing the characterisation above.

(4) Let $C = \bigcap_{\boldsymbol{j}_1,\dots,\boldsymbol{j}_n} G_{\boldsymbol{j}_1,\dots,\boldsymbol{j}_N}$. As C can be biunivocally identified with the set of valid letter exchange maps modulo a shift, return $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d) = \mathbb{Z}^d \times C$ as output.

Example 5.2.2 above corresponds to a simple case in which $G^{(N)} = G^{(1)}$ is a cyclic group, and we derive an abelian subgroup of S_3 corresponding to the valid letter exchange maps. We can use the above procedure to construct examples with more complicated automorphism groups, see Example 5.2.2.

Example We take as alphabet the *quaternion group* $Q_8 = \{e, i, j, k, \bar{e}, \bar{\imath}, \bar{\jmath}, \bar{k}\}$ (see [49] for the multiplication table and basic properties of this group, which is generated by the two elements i and j). With this, we construct a length-3 bijective substitution defined by right multiplication, $x \mapsto (x \cdot i)(x \cdot j)(x \cdot k)$, given in full by:

$$\begin{aligned} e &\mapsto ijk, & \bar{e} &\mapsto \bar{\imath}\bar{\jmath}\bar{k}, \\ i &\mapsto \bar{e}k\bar{\jmath}, & \bar{\imath} &\mapsto e\bar{k}j, \\ j &\mapsto \bar{k}\bar{e}i, & \bar{\jmath} &\mapsto ke\bar{\imath}, \\ k &\mapsto j\bar{\imath}\bar{e}, & \bar{k} &\mapsto \bar{\jmath}ie. \end{aligned}$$

(1) The three permutations obtained from the columns which generate $G^{(1)}$ are:

$$\begin{aligned} R_i &:= (e\, i\, \bar{e}\, \bar{\imath})(j\, \bar{k}\, \bar{\jmath}\, k), \\ R_j &:= (e\, j\, \bar{e}\, \bar{\jmath})(i\, k\, \bar{\imath}\, \bar{k}), \\ R_k &:= (e\, k\, \bar{e}\, \bar{k})(j\, i\, \bar{\jmath}\, \bar{\imath}). \end{aligned}$$

Thus, the substitution θ^3 has as columns $R_{xyz}(g) = g \cdot xyz$ with $x, y, z \in \{i, j, k\}$; in particular, since $jik = e$, θ^3 must have an identity column.

(2) By direct computation, $G^{(n)} = G^{(1)} \cong Q_8$ for all n, making the substitution primitive (as Q_8 acts transitively on itself in an obvious way). Also, since $G^{(3)} = G^{(1)}$, this group is the right Cayley embedding of Q_8 into S_8.

(3) By applying the above algorithm, we obtain that the group of letter exchange maps is generated by the following two permutations:

$$\pi_0 := (e\, i\, \bar{e}\, \bar{\imath})(j\, k\, \bar{\jmath}\, \bar{k}),$$
$$\pi_1 := (e\, j\, \bar{e}\, \bar{\jmath})(i\, \bar{k}\, \bar{\imath}\, k).$$

We can verify that these permutations generate the *left* Cayley embedding of Q_8 into S_8. Alternatively, if we consider the transposition $\nu = (k\, \bar{k})$, we can use Fact 5.7 above to see that $\pi_0 = \nu R_i \nu^{-1}$ and $\pi_1 = \nu R_j \nu^{-1}$, which in turn implies that the group generated by π_0 and π_1 is conjugate to the group generated by R_i and R_j, the latter being isomorphic to Q_8. This shows that $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}) \cong \mathbb{Z} \times Q_8$.

It is well known that automorphism groups of aperiodic minimal one-dimensional subshifts are virtually $\mathbb{Z}$. The following result gives a full converse for shifts generated by bijective substitutions.

Theorem 5.8 ([33, Thm. 3.6]) *For any finite group G, there exists an explicit primitive, bijective substitution θ, on an alphabet on $|G|$ letters, such that* $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}) \cong \mathbb{Z} \times G$. □

The proof, which may be consulted in [33], follows a similar schema to the analysis done in Example 5.2.2 above. In [41, Sec. 4.1], it was shown that the number of letters needed in Theorem 5.8 is actually a tight lower bound. Below, we actually prove something stronger.

Theorem 5.9 *Let θ be an aperiodic, primitive, bijective substitution on the alphabet $\mathcal{A}$. If* $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}) \cong \mathbb{Z} \times G$*, then G must act freely on $\mathcal{A}$, and the order of G has to divide $|\mathcal{A}|$.*

Proof. As seen in [41, Sec. 4.1], if we replace θ with a suitable power, we may ensure that the word $\theta^q(a)$ starts with a and contains every other symbol, for all $a \in \mathcal{A}$. Thus, for any $\pi \in S_n$, the equality $\pi(a) = b$ implies $\pi(\theta^q(a)) = \theta^q(b)$, which in turn determines the images of every symbol in the alphabet; the bound $|G| \leq |\mathcal{A}|$ follows from here.

Note as well that, since θ is bijective, if $\pi(a) \neq a$, then $\pi(c) \neq c$ for every $c \in \mathcal{A}$ as the words $\theta^q(a)$ and $\theta^q(b)$ are either equal or differ at every position. This implies that if π has any fixed point then it must be the identity, i.e. that, if we identify G with the corresponding group of permutations over $\mathcal{A}$, the action of G on the alphabet is free. Equivalently, the stabilizer $\mathrm{Stab}(c)$ of any $c \in \mathcal{A}$ is the trivial subgroup.

The elements of G commute with every column of θ^q. Due to primitivity, there always exists a column $\theta^* = \theta_{\boldsymbol{j}_1} \circ \cdots \circ \theta_{\boldsymbol{j}_q}$ which maps this a to any desired $c \in \mathcal{A}$. Since θ^* commutes with every $\pi \in G$ (i.e., it is an equivariant bijection for the action of G on $\mathcal{A}$), we have that $\mathrm{Orb}(c) = \theta^*[\mathrm{Orb}(a)]$, i.e. the orbit of c under G is necessarily the image of the orbit of a under θ^*.

Thus, every orbit is a set of the same cardinality. This means that G induces a partition of $\mathcal{A}$ into disjoint orbits of the same cardinality ℓ, which then must divide $|\mathcal{A}|$. By the freeness of the group action and the orbit-stabilizer theorem, we have $|G| = |\mathrm{Orb}(a)| \cdot |\mathrm{Stab}(a)| = \ell$, and thus $|G|$ divides $|\mathcal{A}|$.

Remark It follows from Theorem 5.9 that the substitution in Example 5.2.2 is a minimal one in the sense that for one to get a Q-extension in $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z})$, one needs at least eight letters.

Remark At no point in the proof of Theorem 5.8 found in [33] nor in Theorem 5.9 above the fact that the substitution was one-dimensional is actually used. Thus, since Theorem 5.5 is known to be valid for general rectangular substitutions, the two theorems above must be valid in this more general setting as well, provided that the substitution is aperiodic in $\mathbb{Z}^d$, which one can always guarantee; see Propositions 5.4 and 5.21.

Corollary 5.10 *For any finite group G, there exists an explicit primitive and bijective d-dimensional rectangular substitution θ, on an alphabet of $|G|$ letters, such that* $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d) \cong \mathbb{Z}^d \times G$*. Furthermore, this is the least possible alphabet size: for any bijective, primitive and aperiodic d-dimensional rectangular substitution θ on the alphabet $\mathcal{A}$, if* $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d) \cong \mathbb{Z}^d \times G$*, then G acts freely on $\mathcal{A}$, and $|G|$ divides $|\mathcal{A}|$.* □

5.3. Extended and reversing symmetries of substitution shifts

5.3.1. One-dimensional shifts

Since the automorphism group ("symmetry group") does not cover everything that can be thought of as a symmetry (in the geometric sense of the word) we introduce the notion of the **reversing symmetry group**; see [13] for a detailed exposition. We will exclusively look at shift spaces X_θ which are given by a bijective, primitive substitution θ and we will exploit this additional structure in determining the reversing symmetry group for this class.

Once again, we restate the definition of the extended (reversing) symmetry group for ease of access. This group is given by:

$$\mathrm{Sym}(X, \mathbb{Z}^d) := \mathrm{norm}_{\mathrm{Homeo}(X)}(\langle \sigma \rangle) = \{H \in \mathrm{Homeo}(X) \ : \ H\langle \sigma \rangle = \langle \sigma \rangle H\}$$

where $\langle \sigma \rangle$ is the group generated by the shift. In the case where the shift space is one-dimensional, we call $\mathrm{Sym}(X, \mathbb{Z})$ the **reversing symmetry group**, as it can also be described as:

$$\mathrm{Sym}(X, \mathbb{Z}) = \{H \in \mathrm{Homeo}(X) \ : \ H \circ \sigma = \sigma^{\pm 1} \circ H\},$$

and thus any element in the reversing symmetry group that is not an automorphism "reverses the direction" of shift maps, hence the name. Any such element is called a **reversor** or a **reversing symmetry**. Once again, Theorem 3.10 gives a Curtis–Hedlund–Lyndon-type characterisation of reversing and extended symmetries, which incorporates the mirroring component ($\mathrm{GL}(d, \mathbb{Z})$-component in higher dimensions); this is further discussed in [13].

In what follows, we investigate the effect of a reversor f on inflated words. Given a bijective substitution $\theta\colon \mathcal{A} \to \mathcal{A}^L, \theta := \theta_0\theta_1 \cdots \theta_{L-1}$, the mirroring operation m acts on the columns of θ via $m(\theta(a)) = \theta_{L-1}(a) \cdots \theta_2(a)\theta_0(a)$. We may extend this to infinite configurations over $\mathbb{Z}$ in two non-equivalent ways, given by $m(x)_k = x_{-k}$ and $m'(x)_k = x_{1-k}$, respectively; we shall refer to both as basic mirroring maps.

Proposition 5.11 *Let θ be an aperiodic, primitive, bijective substitution. Then, any reversor is a composition of a letter exchange map $\pi \in S_n$, where $n = |\mathcal{A}|$, a shift map σ^k and one of the two basic mirroring maps m or m' (depending only on whether the substitution has odd or even length, respectively).*

See [13, Prop. 1] and Theorem 5.5. This result, while desirable, is not immediately obvious (and can indeed be false for non-bijective substitutions, which may have reversors whose local functions have positive radius), and thus we show this result as a consequence of bijectivity.

Proof. Suppose $f\colon \mathsf{X}_\theta \to \mathsf{X}_\theta$ is a reversor of positive radius $r \geq 1$, i.e., $x|_{[-r,r]} = y|_{[-r,r]}$ implies that one has $f(x)_0 = f(y)_0$. There is some power $k \geq 1$ such that the words $\theta^k(a)$ of length L^k are longer than the local window of f, which has length $2r+1$ (say, $k = \lceil \log(2r+1)/\log(L) \rceil$). Any point of X_θ is a concatenation of words of the form $\theta^k(a), a \in \mathcal{A}$, which is unique up to a shift because of aperiodicity; see [94]. In particular, if we choose a fixed $x \in \mathsf{X}_\theta$ and let $y = f(x)$, both points have such a decomposition.

Now, suppose that the value $L^k = 2\ell + 1$ is odd (the case where L is even is dealt with similarly). By composing f with an appropriate shift map (say $\tilde{f} = f \circ \sigma^h$), we can ensure that the central word $\theta^k(a)$ in the aforementioned decomposition has support $[-\ell, \ell]$ for both x and y (note that we employ the uniqueness of the decomposition here, to avoid ambiguity in the chosen h). Since $L^k = 2\ell + 1 \geq 2r + 1$, we must have $\ell \geq r$, and thus y_0 is entirely determined by $x|_{[-\ell,\ell]}$, which is a substitutive word $\theta^k(a)$. But, since θ is bijective, this word is in turn completely determined by its central symbol x_0.

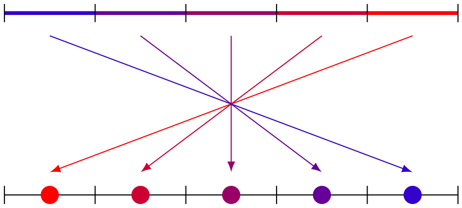


Figure 5.1: A reversor f establishes a 1-1 correspondence between words $\theta^k(a)$ in a point x and its image $f(x)$.

A similar argument shows that, for any $n \in \mathbb{Z}$, if $n \in mL^k + [-\ell, \ell]$, then y_n depends only on the word $x|_{-mL^k+[-\ell,\ell]}$, which contains (and is thus entirely determined by) x_{-n}. Since any point in X_θ is transitive, $\tilde{f}$ is entirely determined by the points x and y, and thus, $\tilde{f}$ is a map of radius 0. Equivalently, for some bijection $\pi\colon \mathcal{A} \to \mathcal{A}$, we have $\tilde{f}(x)_{-n} = \pi(x_n)$, that is, $\tilde{f} = f \circ \sigma_h = \pi \circ m$ (identifying π with the letter exchange map $\mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$). We conclude that f is a composition of a letter exchange map, a mirroring map and a shift map.

Remark With some care, it can be shown that the same argument applies in the higher-dimensional case, where an element of the normalizer is a composition of a letter exchange map, a map of the form $f(x)_{\boldsymbol{n}} = x_{A\boldsymbol{n}}$, with A a linear map from the hyperoctahedral group (see Theorem 5.16, below), and a shift map; see [13, Prop. 3] for a more general formulation.

This result leads to the following criterion for the existence of a reversor in terms of the columns θ_i.

Proposition 5.12 *Let θ be an aperiodic, primitive, bijective substitution θ of length L on a finite alphabet $\mathcal{A}$ of n letters. Suppose that there exists a letter-exchange map $\pi \in S_n$, $\pi : \mathcal{A} \to \mathcal{A}$ which gives rise to a reversing symmetry. Then one has*

$$\pi^{-1} \circ \theta_i \circ \theta_j^{-1} \circ \pi = \theta_{L-(i+1)} \circ \theta_{L-(j+1)}^{-1} \tag{5.2}$$

for all $0 \leq i, j \leq L-1$, where θ_i is the i-th column of θ seen as an element of S_n.

Proof. Let $a \in \mathcal{A}$. Let m be the mirroring operation and suppose that there exists $\pi \in S_n$ such that $m \circ \pi$ extends to a reversor $f \in \mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z})$. One then has

$$\theta(a) = \theta_0(a) \cdots \theta_{L-1}(a) \xmapsto{m} \theta_{L-1}(a) \cdots \theta_0(a) \xmapsto{\pi} \pi \circ \theta_{L-1}(a) \cdots \pi \circ \theta_0(a).$$

Since Proposition 5.11 guarantees that this must result to mapping substituted words to substituted words, one gets

$$\pi \circ \theta_{L-1}(a) \cdots \pi \circ \theta_0(a) = \theta_0(b) \cdots \theta_{L-1}(b) = \theta_0 \circ \tau(a) \cdots \theta_{L-1} \circ \tau(a), \tag{5.3}$$

where the permutation τ describes precisely this induced shuffling of inflation words. This yields

$$\tau = \theta_j^{-1} \circ \pi \circ \theta_{L-(j+1)}$$

for all $0 \leq j \leq L-1$. Equating the corresponding right hand-sides for some pair i, j yields Eq. (5.2). The claim follows since this must hold for all $0 \leq i, j \leq L-1$.

Theorem 5.13 *Let θ be as in Proposition 5.12. Suppose further that $\theta_i = \theta_{L-(i+1)} = \mathrm{id}$ for some $0 \leq i \leq L-1$. Then, given a permutation (letter exchange map) $\pi \in S_n, \pi\colon \mathcal{A} \to \mathcal{A}$, the following are equivalent:*

(i) The letter exchange map π gives rise to a reversing symmetry $f \in \mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}) \setminus \mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z})$ given by either $f(x)_n = \pi(x_{-n})$ or $f(x)_n = \pi(x_{1-n})$.

(ii) The permutation π satisfies the system of equations

$$\pi^{-1} \circ \theta_i \circ \pi = \theta_{L-(i+1)} \tag{5.4}$$

for all $0 \leq i \leq L-1$.

(iii) There exist $\kappa_0, \kappa_1, \ldots, \kappa_{L-1} \in S_n$, where each κ_i satisfies $\kappa_i^{-1} \circ \theta_i \circ \kappa_i = \theta_{L-(i+1)}$, such that the following intersection of cosets is non-empty:

$$K = \bigcap_{i=0}^{L-1} \mathrm{cent}_{S_n}(\theta_i)\kappa_i, \tag{5.5}$$

and $\pi \in K$.

Proof. It is clear that Eq. (5.4) implies Eq. (5.2). Note that it is sufficient to satisfy Eq. (5.2) for $j = i + 1 \bmod L$ as any term can be obtained by multiplying sufficient numbers of succeeding terms. Under the extra assumption that there exist a column pair which is the identity, Eq. (5.2) simplifies to Eq. (5.4). This shows that (i) $\implies$ (ii).

For the other direction, we show that if Eq. (5.4) is satisfied at by the level-1 inflation words, then these sets of equations must also be fulfilled by any power θ^k of θ. Remember that, from any arbitrary bijective substitution θ, we may derive another bijective substitution θ' that satisfies the additional condition of having two identity columns in opposing positions by choosing $k = \operatorname{lcm}(|\theta_0|, |\theta_{L-1}|)$ and replacing θ by its k-th power, $\theta' := \theta^k$. This makes no difference when studying $\operatorname{Sym}(\mathsf{X}_\theta, \mathbb{Z})$, because θ and θ^k define the same subshift and the group of reversing symmetries is a property of the hull.

First, we prove an important property of the columns of powers. Fix a power $k \in \mathbb{N}$ and pick a column θ_i of θ^k, where $0 \le i \le L^k - 1$. One then has $\theta_i = \theta_{i_0} \cdots \theta_{i_{k-1}}$ where $i_0 i_1 \cdots i_{k-1}$ is the L-adic expansion of i and θ_{i_ℓ} are columns of the level-1 substitution θ.

The corresponding L-adic expansion of $L^k - (i+1)$ is then given by

$$L^k - (i+1) = (L - (i_0 + 1)) \cdots (L - (i_{k-1} + 1)).$$

This can easily be shown via the following direct computation

$$\sum_{j=0}^{k-1} (L - (i_j + 1)) L^j = \sum_{j=0}^{k-1} (L^{j+1} - L^j) - \sum_{j=0}^{k-1} i_j L^j = L^k - (i+1).$$

This implies that if one considers the corresponding column $\theta_{L^k-(i+1)}$ one gets that

$$\theta_{L^k-(i+1)} = \theta_{L-(i_0+1)} \cdots \theta_{L-(i_{k-1}+1)}. \tag{5.6}$$

This has two consequences. First, if θ has an identity column pair, then all powers of θ admit at least one identity column pair. For each power k one just needs to choose θ_j with $j = iii \cdots i$, which implies $\theta_j = \theta_i^k = \mathrm{id}$. By Eq. (5.6), we also get that $\theta_{L^k-(j+1)} = (\theta_{L-(i+1)})^k = \mathrm{id}$. In fact, θ^k contains at least 2^{k-1} pairs of identity columns.

Second, this property allows one to prove that if θ satisfies the system of equations in Eq. (5.4), then it is satisfied at all levels, i.e., by all powers of θ. To this end, choose $0 \le i \le L^k - 1$ with L-adic expansion $i_0 i_1 \cdots i_{k-1}$. From Eq. (5.4) one then obtains

$$\begin{aligned} \pi^{-1} \circ \theta_i \circ \pi = \pi^{-1} \circ \theta_{i_0} \cdots \theta_{i_{k-1}} \circ \pi &= \pi^{-1} \circ \theta_{i_0} \pi \pi^{-1} \cdots \pi \pi^{-1} \theta_{i_{k-1}} \pi \\ &= \theta_{L-(i_0+1)} \cdots \theta_{L-(i_{k-1}+1)} = \theta_{L^k-(i+1)}. \end{aligned}$$

Since i is chosen arbitrarily and π induces a permutation of the substituted words at all levels, this means it extends to a map $f = \sigma_n \circ m \circ \pi \colon \mathsf{X}_\theta \to \mathsf{X}_\theta$, which by Proposition 5.11 is a reversor. This shows (ii) $\implies$ (i).

To prove the remaining equivalences, note that if $\pi_1, \pi_2 \in S_n$ are two permutations satisfying the equality $\pi^{-1} \circ \theta_i \circ \pi = \theta_{L-(i+1)}$, then we have:

$$\pi_1 \circ \theta_{L-(i+1)} \circ \pi_1^{-1} = \theta_i \implies (\pi_2 \circ \pi_1^{-1})^{-1} \circ \theta_i \circ (\pi_2 \circ \pi_1^{-1}) = \theta_i,$$

that is, $(\pi_2 \circ \pi_1^{-1}) \in \mathrm{cent}_{S_n}(\theta_i)$. As a consequence, π_1 belongs to the right coset $\mathrm{cent}_{S_n}(\theta_i)\pi_2$ for any choice of π_1, π_2, and, since right cosets are either equal or disjoint, this means that all solutions of Eq. (5.4), for a fixed i, lie in the same right coset of $\mathrm{cent}_{S_n}(\theta_i)$. Reciprocally, if π satisfies Eq. (5.4) and $\gamma \in \mathrm{cent}_{S_n}(\theta_i)$, it is easy to verify that $\gamma \circ \pi$ satisfies Eq. (5.4) as well. Thus, the set of solutions of this equation is either empty or the aforementioned uniquely defined right coset.

Thus, suppose that π satisfies Eq. (5.4) for all $0 \leq i \leq L-1$. The set of solutions for each i equals the unique coset $\mathrm{cent}_{S_n}(\theta_i)\pi$, and thus the set of all permutations that satisfy Eq. (5.4) for all i is exactly the intersection of all these cosets, i.e. $\bigcap_{i=0}^{L-1} \mathrm{cent}_{S_n}(\theta_i)\pi$. Taking $\kappa_i = \pi$ for all i, we see that this is exactly the set K from (5.5). Evidently, π belongs to this intersection, and so we conclude that (ii) $\implies$ (iii).

As stated before, our choice of κ_i ensures that the set $\mathrm{cent}_{S_n}(\theta_i)\kappa_i$ is exactly the set of solutions of Eq. (5.4) for a given i; thus, any permutation π that satisfies all of these equalities must be in all of these cosets and thus in the intersection (5.5), which is therefore non-empty. This shows that (iii) $\implies$ (ii), concluding the proof.

The following general criterion on when a letter-exchange map generates a reversor is given in [13].

Lemma 5.14 ([13, Lem. 2]) *Let θ be a primitive constant-length substitution of height* 1 *and column number c_θ. Suppose that θ is strongly injective. Then, a permutation $\pi \colon \mathcal{A} \to \mathcal{A}$ generates a reversor $f \in \mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z})$ if and only if*

1. $ab \in \mathcal{L}_2(\mathsf{X}_\theta) \implies \pi(ba) \in \mathcal{L}_2(\mathsf{X}_\theta)$
2. $(\pi \circ \theta^{c_\theta !})(ab) = (\theta^{c_\theta !} \circ \pi)(ba)$

for each $ab \in \mathcal{L}_2(\mathsf{X}_\theta)$. □

For a primitive, aperiodic and bijective θ, one has $c_\theta = |\mathcal{A}|$. Moreover, θ is always strongly injective. Note that Theorem 5.13 implies conditions **(1)** and **(2)** in Lemma 5.14. The first one immediately follows from primitivity, and the fact that any legal word ab appears in some level-n superword, which is sent to another level-n superword by $\pi \circ m$, which guarantees the legality of $\pi(ba)$. The second follows from the fact that π is compatible with superwords of all levels. In fact, one has $\big(\pi \circ \theta^n\big)(ab) = \big(\theta^n \circ \pi\big)(ba)$ for all $n \in \mathbb{N}$.

Remark It is a known fact from group theory that, if $g_1, \ldots, g_r$ are elements of a group G and $H_1, \ldots, H_r$ are subgroups of this group, the intersection of cosets $\bigcap_{i=1}^{r} g_i H_i$ is either empty or a coset of $\bigcap_{i=1}^{r} H_i$. In this case, the latter intersection is exactly the group of non-trivial automorphisms modulo a shift (letter exchanges), and thus, if there exist non-trivial reversing symmetries, these must all belong to a single coset of the group of valid letter exchanges. This is consistent with the fact that $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z})$ is at most an index 2 group extension of $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z})$.

Item (3) in Theorem 5.13 provides an explicit algorithm to compute the group of permutations π which define extended symmetries, which is a counterpart to that in Section 5.2.2 for automorphisms. As stated previously, the centralizers $\mathrm{cent}_{S_n}(\theta_j)$ can be computed for each column using Fact 5.7, and thus the problem reduces to obtaining a suitable candidate for each κ_i, which once again can be done by an application of Fact 5.7. The algorithm is as follows:

Algorithm. Assuming that θ is a primitive, bijective, aperiodic substitution, the following algorithm computes the set K of permutations that induce reversors, which determines $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z})$.

- **Input:** θ is a length-L bijective substitution, represented either as a function or a set of columns.
- **Output:** A (finite) set of permutations K, either empty or a coset of the group C computed by the previous algorithm, so that $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z})/\langle\sigma\rangle \cong C \cup K$ (i.e. $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}) \cong \mathbb{Z} \rtimes_\varphi (C \cup K)$, with $\varphi(g, n) = n$ if $g \in C$, and $-n$ if $g \in K$).

(1) Let N be the least positive integer which ensures that two opposite columns of θ^N are the identity map. This can be computed as:

$$N = \text{mín}\left\{\mathrm{lcm}(\mathrm{ord}(\theta_i), \mathrm{ord}(\theta_{L-(i+1)})) \ : \ 0 \leq i \leq N/2\right\}.$$

(2) For each $0 \leq i \leq N/2$, compute κ_i via the following subroutine:

(2.i) If θ_i and $\theta_{L-(i+1)}$ are non-conjugate (i.e., their cycle decomposition has a different number of cycles of some length), stop the algorithm, as reversors do not exist (see Theorem 5.13).

(2.ii) Sort the cycles from the disjoint cycle decomposition of θ_i by increasing order of length. Using this as a basis, by appropriately sorting the elements of each cycle in this decomposition, define a total order relation $<$ on $\mathcal{A}$, given by, say, $a_1 < \cdots < a_n$, such that all of the elements of a given cycle come before the elements of the following cycle, in the sorting by left. Do the same for $\theta_{L-(i+1)}$, defining a corresponding total order $<'$ given by $b_1 <' \cdots <' b_n$. This ensures that there are cycle decompositions of both permutations such that the corresponding cycles, ordered from left to right, have the same length, as follows:

$$\theta_i = (a_1 \, \dots \, a_j)(a_{j+1} \, \dots \, a_{j'}) \cdots (a_{j''+1} \, \dots \, a_n),$$
$$\theta_{L-(j+1)} = (b_1 \, \dots \, b_j)(b_{j+1} \, \dots \, b_{j'}) \cdots (b_{j''+1} \, \dots \, b_n),$$

with $1 \leq j \leq j' \leq \ldots \leq j'' \leq n$.

(2.iii) Define:

$$\kappa_i = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{pmatrix}, \qquad \kappa_{L-(i+1)} = \kappa_i^{-1}.$$

(3) Compute each centralizer $C^{(i)} = \mathrm{cent}_{S_n}(\theta_i)$, using the same procedure as in the computation of $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z})$.

(4) Return $K = \bigcap_{i=1}^{N} C^{(i)}\kappa_i$. Any element of K induces a reversor; if K is empty, reversors do not exist.

Any programming environment with suitable data structures (e.g. computer algebra systems such as `Sagemath`® or `Mathematica`®) is amenable to the implementation of this algorithm, providing effective procedures to entirely characterise the groups $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z})$ and $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z})$ from a suitable description of the substitution θ, e.g. using a dictionary.

Example Going back to Example 5.2.2, we may apply the previous algorithm to determine whether reversors for this substitution do exist. Following the steps of Algorithm 2, we obtain:

(1) For the algorithm to work properly, we need two columns in opposite positions to be identity columns. Since every element in the quaternion group Q_8 has order 4, we may just take $N = 4$ (and indeed, inspection shows that this is the smallest value of N that satisfies this property).

(2) It is not hard to see that the columns of θ^4 are, in order:

$$R_{i^4}, R_{i^3 j}, R_{i^3 k}, R_{i^2 ji}, \dots, R_{k^2 jk}, R_{k^3 i}, R_{k^3 j} R_{k^4},$$

and thus, due to the nature of the elements of Q_8 (namely, that the mapping that sends i and j to any two of the three elements $\{i, j, k\}$ is a group automorphism), opposite columns are conjugate.

We need to find mappings κ_r such that the r-th column from left to right is conjugate to the corresponding column from right to left under k. For example, the second and penultimate column are given by:

$$\begin{aligned} R_{i^3 j} &= R_{\bar{k}} = (e\,\bar{k}\,\bar{e}\,k)(i\,j\,\bar{\imath}\,\bar{\jmath}), \\ R_{k^3 j} &= R_i = (e\,i\,\bar{e}\,\bar{\imath})(j\,\bar{k}\,\bar{\jmath}\,k). \end{aligned}$$

Using Fact 5.7, we see that if κ_1 is a permutation that maps $R_{i^3 j}$ to $R_{k^3 j}$ via conjugation, choosing the images of one element of the first cycle and one of the second is enough to determine the whole permutation. If $\kappa_1(e) = k$ and $\kappa_1(i) = e$, then it must map the following elements of each cycle of $R_{i^3 j}$ to the following elements of the corresponding cycle in $R_{k^3 j}$, and thus we obtain:

$$\begin{aligned} \kappa_1 &= \begin{pmatrix} e & i & j & k & \bar{e} & \bar{\imath} & \bar{\jmath} & \bar{k} \\ \bar{\jmath} & i & \bar{e} & \bar{k} & j & \bar{\imath} & e & k \end{pmatrix} \\ &= (e\,\bar{\jmath})(j\,\bar{e})(\bar{k}\,k). \end{aligned}$$

Thus, any element of the coset $\mathrm{cent}_{S_8}(R_{i^3 j})\kappa_1$ in S_8 maps the second column to the penultimate one by conjugation. Note that the corresponding step of Algorithm 2 above actually returns a different permutation, $\kappa_1' = (\bar{k}\,i\,j)(k\,\bar{\imath}\,\bar{\jmath})$, but direct computation shows that κ_1 and κ_1' belong to the same coset of $\mathrm{cent}_{S_8}(R_{i^3 j})$ and thus the algorithm proceeds in the same way for either; we choose κ_1 instead of κ_1' for mere convenience. After this, we repeat the same procedure for the remaining 40 pairs of columns (including the center, which is paired with itself) and compute the intersection of the obtained cosets.

(3) We note that the computed permutation κ_1 appears in every coset $\mathrm{cent}_{S_8}((\theta^4)_r)\kappa_r$, and thus the intersection of all cosets involved equals a right coset of the left Cayley embedding of Q_8 in S_8, which must equal $K = L(Q_8)\kappa_1$. It can be verified from computation that the union $L(Q_8) \cup K$ of this embedding and the corresponding coset is also a subgroup of S_8.

(4) Thus, every element of $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z})$ is associated with a letter swap from the subgroup $G = L(Q_8) \cup K$ of S_8, with reversors corresponding to elements of $K = G \setminus Q_8$. Note that this group has order 16. Besides, $\{R_e, \kappa_1\}$ is an order 2 subgroup of G with trivial

intersection with Q_8, which is normal in G due to being of index 2; thus, this group has a natural semidirect product structure as $G \cong Q_8 \rtimes C_2$.

Computation aided with computer algebra software shows that this group G has 15 subgroups and 7 different conjugacy classes. The only group of order 16 with both properties is the *semidihedral group*, SD_{16}. Thus, we obtain a complete description of $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z})$ as the semidirect product $\mathbb{Z} \rtimes SD_{16} \cong (\mathbb{Z} \times Q_8) \rtimes C_2$.

5.3.2. Higher-dimensional subshifts

Now, we turn our attention to the situation in higher dimensions, and thus we will speak of "extended symmetries" instead of reversing symmetries, once again.

Similar to automorphisms, there is a direct generalisation of the characterisation of extended symmetries from Proposition 5.12 and the subsequent theorem to the higher-dimensional setting, which is given by the following.

Proposition 5.15 *Let θ be an aperiodic, primitive, bijective, block substitution in $\mathbb{Z}^d$. Then any extended symmetry $f \in \mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d) \setminus \mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ must be (up to a shift) a composition of a letter exchange map and a rearrangement function f_A given by $f_A(x)_{\boldsymbol{n}} = x_{A\boldsymbol{n}}$, where $A \in \mathrm{GL}_d(\mathbb{Z})$, with $A \neq \mathbb{I}$.* □

For shifts generated by bijective rectangular substitutions one has the following restriction on the linear component A of an extended symmetry f.

Theorem 5.16 ([22, Thm. 18]) *Let θ an aperiodic, primitive, bijective rectangular substitution in $\mathbb{Z}^d$. One then has*

$$\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)/\,\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d) \cong P \leq W_d,$$

where $W_d \cong C_2^d \rtimes S_d$ is the d-dimensional hyperoctahedral group, which represents the symmetries of the d-dimensional cube. □

With this, one can show that all extended symmetries of such subshifts are of finite order. The proof of the following result is patterned from [12, Prop. 2], which deals with the order of reversors of an automorphism h of a general dynamical system with $\mathrm{ord}(h) = \infty$; compare [47].

Proposition 5.17 *Let X_θ be the same as above with automorphism group $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d) = \mathbb{Z}^d \times G$. Let $f \in \mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d) \setminus \mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ be an extended symmetry, whose associated matrix is $A \in W_d \backslash \mathbb{I}$. Then $\mathrm{ord}(f)$ divides $\mathrm{ord}(A) \cdot |G|$. Moreover, $\mathrm{ord}(f) \leq 2|G| \cdot \mathrm{máx}\,\{\mathrm{ord}(\tau) \mid \tau \in S_d\}$.*

Proof. Under the given assumptions, $f \circ \sigma_{\boldsymbol{m}} \circ f^{-1} = \sigma_{A\boldsymbol{m}}$ holds for all $\boldsymbol{m} \in \mathbb{Z}^d$, which yields

$$f^\ell \circ \sigma_{\boldsymbol{m}} \circ f^{-\ell} = \sigma_{A^\ell \boldsymbol{m}} \tag{5.7}$$

$$f \circ \sigma_{n\boldsymbol{m}} \circ f^{-1} = \sigma_{nA\boldsymbol{m}} \tag{5.8}$$

for all $\ell, n \in \mathbb{N}$. Choosing $\ell = \mathrm{ord}(A)$, Eq. (5.7) gives $f^{\mathrm{ord}(A)} \in \mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$. From Theorem 5.5, $f^{\mathrm{ord}(A)} = \sigma_{\boldsymbol{p}} \circ \pi$, for some $\boldsymbol{p} \in \mathbb{Z}^d$ and letter-exchange map π. From the direct product structure of the automorphism group, one has $\sigma_{\boldsymbol{p}} \circ \pi = \pi \circ \sigma_{\boldsymbol{p}}$, which implies $f^{\mathrm{ord}(A)\cdot|G|} = \sigma_{|G|\boldsymbol{p}} \circ \pi^{|G|} = \sigma_{|G|\boldsymbol{p}}$. Using the two equations above, one gets $f^{\mathrm{ord}(A)\cdot|G|} = \sigma_{|G|A^{\ell}(\boldsymbol{p})}$ for all $\ell \in \mathbb{N}$. Since f is an extended symmetry, $A \neq \mathbb{I}$. Next we show that $\boldsymbol{p}$ cannot be an eigenvector of A.

Suppose $A\boldsymbol{p} = \boldsymbol{p}$ with $\boldsymbol{p} \neq \boldsymbol{0}$. Note that $f^{-\mathrm{ord}(A)|G|} = \sigma_{-|G|\boldsymbol{p}}$. From Eqs. (5.7) and (5.8), one also has $f^{-1} \circ \sigma_{|G|A^{-1}\boldsymbol{p}} \circ f = \sigma_{-|G|\boldsymbol{p}}$, which implies $A^{-1}\boldsymbol{p} = -\boldsymbol{p}$, contradicting the assumption on $\boldsymbol{p}$. Since $\mathrm{ord}(\sigma_{\boldsymbol{p}}) = \infty$, this forces $\boldsymbol{p} = \boldsymbol{0}$ and hence $f^{\mathrm{ord}(A)\cdot|G|} = \mathrm{id}$ from which the first claim is immediate. The upper bound for the order follows from the upper bound for the order of the elements of the hyperoctahedral group W_d; see [2].

Due to the fact that $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)$ is (possibly) a larger extension of $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ (that is, the corresponding quotient can have up to $2^d d! - 1$ non-trivial elements instead of just one), we would end up with a much larger number of equations of the form of Eq. (5.2), one for each element of the hyperoctahedral group W_d except the identity. This leads us to another problem of different nature: if the rectangle R, which is the support of the level-1 supertiles of θ, is not a cube in $\mathbb{Z}^d$, some symmetries from W_d may not be compatible with R, i.e., they may map R to a different rectangle that is not a translation of R, so the corresponding equation does not have a proper meaning (as it may compare an existing column with a non-existent one).

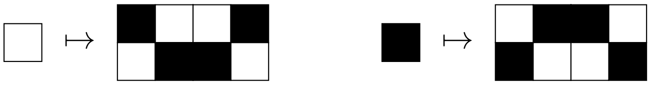


Figure 5.2: A non-square substitution that generates the two-dimensional Thue-Morse hull.

This could be taken as a suggestion that such geometrical symmetries cannot actually happen, imposing further limitations on the quotient $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)/\,\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$. Interestingly, this is not actually the case. For instance, consider the two-dimensional rectangular substitution from Figure 5.2. As the support for this substitution is a 4×2 rectangle, we could guess that this substitution is incompatible with rotational symmetries or reflections along a diagonal axis, which would produce a 2×4 rectangle instead. However, further examination shows that the hull generated by this substitution is actually the same as the hull of the two-dimensional Thue–Morse substitution as seen in e.g. [22], which is compatible with every symmetry from $W_2 = D_4$. Thus, only geometrical considerations are not enough to exclude candidates for extended symmetries.

Fortunately, there is a subcase of particular interest in which this geometrical intuition is actually correct, which involves an arithmetic restriction on the side lengths of the support rectangle R. It turns out that coprimality of the side lengths is a sufficient condition (although it can be weakened even further) to rule out such symmetries, e.g. there are no extended symmetries compatible with rotations when R is a, say, 2×5 rectangle. To be precise:

Theorem 5.18 *Let $\theta\colon \mathcal{A} \to \mathcal{A}^R$ be a bijective rectangular substitution with faithful associated shift action. Suppose that $R = [\boldsymbol{0}, \boldsymbol{L} - \boldsymbol{1}]$ with $\boldsymbol{L} = (L_1, \ldots, L_d)$ (that is, R is a d-dimensional rectangle with side lengths $L_1, L_2, \ldots, L_d$) and that for some indices i, j there is a prime p such that $p \mid L_j$ but $p \nmid L_i$, i.e. L_i and L_j have different sets of prime factors. Let $A \in W_d \leq \mathrm{GL}(d, \mathbb{Z})$ and suppose that A is the underlying matrix associated to an extended symmetry $f \in \mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)$. Then $A_{ij} = A_{ji} = 0$.*

The underlying idea is that, if $A \in W_d$ induces a valid extended symmetry for some substitution θ with support U, we can find another substitution η with support $A \cdot U$ (up to an appropriate translation) such that $\mathsf{X}_\theta = \mathsf{X}_\eta$, and then we use the known factor map from an aperiodic substitutive subshift onto an associated odometer to rule out certain matrices A. Similar exclusion results have been studied by Cortez and Durand [26].

Proof. Let $\varphi \colon \mathsf{X}_\theta \twoheadrightarrow \mathbb{Z}_{L_1} \times \cdots \times \mathbb{Z}_{L_d} = \mathbb{Z}_{\boldsymbol{L}}$ be the standard factor map from the substitutive subshift to the corresponding product of odometers. It is known [13, Thm. 5] that, for any extended symmetry $f \colon \mathsf{X}_\theta \to \mathsf{X}_\theta$ with associated matrix A, there exists $\boldsymbol{k}_f = (k_1, \dots, k_d) \in \mathbb{Z}_{\boldsymbol{L}}$ and a group automorphism $\alpha_f \colon \mathbb{Z}_{\boldsymbol{L}} \to \mathbb{Z}_{\boldsymbol{L}}$ satisfying the following equation:

$$\varphi(f(x)) = \boldsymbol{k}_f + \alpha_f(\varphi(x)), \tag{5.9}$$

where α_f is the unique extension of the map $\boldsymbol{n} \mapsto A\boldsymbol{n}$, defined in the dense subset $\mathbb{Z}^d$, to $\mathbb{Z}_{\boldsymbol{L}}$. In particular, for any $\boldsymbol{n} \in \mathbb{Z}^d$, if $f = \sigma_{\boldsymbol{n}}$ is a shift map, then $\boldsymbol{k}_{\sigma_{\boldsymbol{n}}} = \boldsymbol{n}$ and $\alpha_{\sigma_{\boldsymbol{n}}} = \mathrm{id}_{\mathbb{Z}_{\boldsymbol{L}}}$.

Now, consider the sequence $\boldsymbol{h}_m = L_i^m \boldsymbol{e}_i$, and suppose $A_{ji} = \pm 1$. Equivalently, $A\boldsymbol{e}_i = \pm \boldsymbol{e}_j$, since A is a signed permutation matrix. Without loss of generality, we may assume the sign to be +. One has $L_i^m \xrightarrow{m\to\infty} 0$ in the L_i-adic topology, and thus $\varphi(\sigma_{\boldsymbol{h}_m}(x)) = \boldsymbol{h}_m + \varphi(x) \xrightarrow{m\to\infty} \varphi(x)$, as it does so componentwise. By compactness, we may take a subsequence $\boldsymbol{h}_{\beta(m)}$ such that $\sigma_{\boldsymbol{h}_{\beta(m)}}(x)$ converges to some x^*; then, as the factor map φ is continuous, we have $\varphi(x^*) = \varphi(x)$.

Eq. (5.9) and this last equality imply that $\varphi(f(x)) = \varphi(f(x^*))$ as well. Writing x^* as a limit, we obtain from continuity that

$$\begin{aligned} \varphi(x^*) &= \operatorname*{lím}_{m\to\infty} \varphi(f(\sigma_{\boldsymbol{h}_{\beta(m)}}(x))) = \operatorname*{lím}_{m\to\infty} \varphi(\sigma_{A\boldsymbol{h}_{\beta(m)}}(f(x))) \\ &= \varphi(x) + \operatorname*{lím}_{m\to\infty} A\boldsymbol{h}_{\beta(m)} = \varphi(x) + \operatorname*{lím}_{m\to\infty} L_i^{\beta(m)} A\boldsymbol{e}_i \\ \Longrightarrow \operatorname*{lím}_{m\to\infty} L_i^{\beta(m)} \boldsymbol{e}_j &= \varphi(x^*) - \varphi(x) = \boldsymbol{0}. \end{aligned}$$

The last equality implies that, in the topology of $\mathbb{Z}_{L_j}$, the sequence $L_i^{\beta(m)}$ converges to 0. However, since there is a prime p that divides L_j but not L_i, due to transitivity we must have $L_j \nmid L_i^n$ for all n, as otherwise $p \mid L_i^n$ and thus $p \mid L_i$. Thus, in base L_j, the last digit of $L_i^{\beta(m)}$ is never zero, and thus $L_i^{\beta(m)}$ remains at fixed distance 1 from $\boldsymbol{0}$ (in the L_j-adic metric), contradicting this convergence. Thus, A_{ji} cannot be 1 and must necessarily equal 0. For A_{ij}, the same reasoning applies to f^{-1}. Since A is a signed permutation matrix, $A_{ij} = \pm 1$ would imply $(A^{-1})_{ji} = \pm 1$, again a contradiction.

We now proceed to the generalisation of Theorem 5.13 in higher dimensions. As before, for a block substitution θ, we have $R = \prod_{i=1}^{d} [0, L_i - 1]$, with $L_i \geq 2$ and the expansive map $Q = \mathrm{diag}(L_1, L_2, \dots, L_d)$. Let $A \in W_d \leq \mathrm{GL}(d, \mathbb{Z})$ be a signed permutation matrix. First, we assume that the location of a tile in any supertile is given by the location of its centre. Define the affine map $A^{(1)} \colon R \to R$ via $A^{(1)}(\boldsymbol{i}) = A(\boldsymbol{i} - \boldsymbol{x}_1) + |A|\boldsymbol{x}_1$ where $\boldsymbol{i} \in R$ and $\boldsymbol{x}_1 = Q\boldsymbol{v} - \boldsymbol{v}$ with $\boldsymbol{v} = \frac{1}{2}(1, 1, \dots, 1)^T$. Here, $(|A|)_{ij} = |A_{ij}|$. The vector $|A|\boldsymbol{x}_1$ is the translation needed to shift the centre of the supertile to the origin, which we will need before applying the map A and shifting it back again. We extend $A^{(1)}$ to any level-k supertile by defining the map $A^{(k)} \colon R^{(k)} \to R^{(k)}$ given by

$$A^{(k)}(\boldsymbol{i}) = A(\boldsymbol{i} - \boldsymbol{x}_k) + |A|\boldsymbol{x}_k, \tag{5.10}$$

with $\boldsymbol{i} \in R^{(k)}$ and $\boldsymbol{x}_k = Q^k \boldsymbol{v} - \boldsymbol{v}$. Here $R^{(k)} := \prod_{i=1}^{d} [0, L_i^k - 1]$ is the set of locations of tiles in a level-k supertile.

Example Let θ be a two-dimensional block substitution with $Q = \left(\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right)$ and A be the counterclockwise rotation by 90 degrees, with corresponding matrix $A = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. Consider the level-3 supertile and let $\boldsymbol{i} = (7, 3)^T \in R^{(3)}$, with Q-adic expansion $\boldsymbol{i} \mathrel{\widehat{=}} \boldsymbol{i}_2 \boldsymbol{i}_1 \boldsymbol{i}_0$. Here one has $\boldsymbol{i}_0 = \boldsymbol{i}_1 = \boldsymbol{e}_1 + \boldsymbol{e}_2$ and $\boldsymbol{i}_2 = \boldsymbol{e}_1$. One then gets $A^{(3)}(\boldsymbol{i}) = (4, 7)^T$; see Figure 5.3. One can check that $\sum_{j=0}^{2} Q^j (A^{(1)}(\boldsymbol{i}_j)) = A^{(3)}(\boldsymbol{i})$.

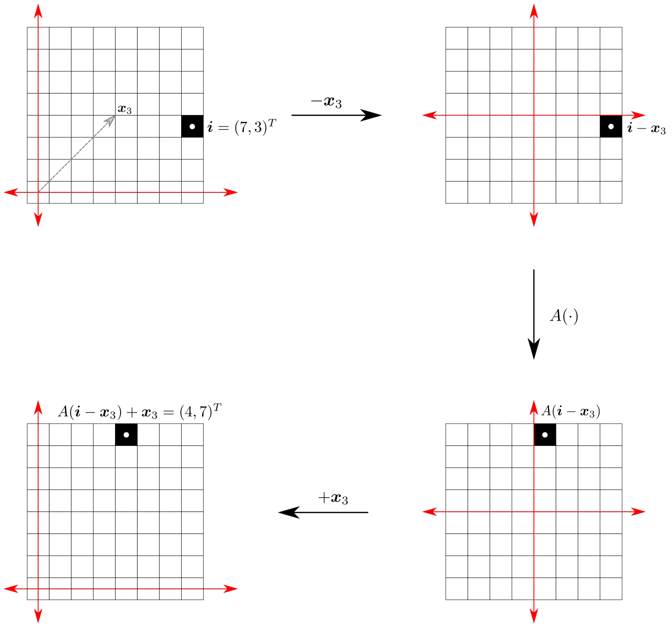

$\boldsymbol{x}_3$
$\boldsymbol{i} = (7,3)^T$
$-\boldsymbol{x}_3$
$\boldsymbol{i} - \boldsymbol{x}_3$
$A(\cdot)$
$A(\boldsymbol{i} - \boldsymbol{x}_3) + \boldsymbol{x}_3 = (4,7)^T$
$+\boldsymbol{x}_3$
$A(\boldsymbol{i} - \boldsymbol{x}_3)$

Figure 5.3: The transformation of a marked level-3 location set $R^{(3)}$ under the map $A^{(3)}$.

The following result is the analogue of Theorem 5.13 in $\mathbb{Z}^d$.

Theorem 5.19 *Let θ be an aperiodic, primitive, bijective block substitution $\theta\colon \mathcal{A} \to \mathcal{A}^R$. Let W_d be the d-dimensional hyperoctahedral group and let $A \in W_d$. Suppose there exists $\boldsymbol{\ell} \in R$ such that $\theta_{\boldsymbol{\ell}'} = \mathrm{id}$ for all $\boldsymbol{\ell}' \in \mathrm{Orb}_A(\boldsymbol{\ell})$. Assume further that $[A, Q] = 0$ and $|A|\boldsymbol{x}_1 = \boldsymbol{x}_1$. Then π, together with A, gives rise to an extended symmetry $f \in \mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)$ if and only if*

$$\pi^{-1} \circ \theta_{\boldsymbol{i}} \circ \pi = \theta_{A^{(1)}(\boldsymbol{i})} \tag{5.11}$$

for all $\boldsymbol{i} \in R$.

Proof. Most parts of the proof mimics those of the proof of Theorem 5.13, where one replaces the mirroring operation m with a more general map $A \in W_d$. One then gets an analogous system of equations, as in those coming from Eq. (5.3). Using this, one can show the necessity direction.

To prove sufficiency, we show that if Eq. (5.11) is satisfied for all $\boldsymbol{i} \in R$, then it also holds for all positions in any level-k supertile. Let $\boldsymbol{i} \in R^{(k)}$, which admits the unique Q-adic expansion given by $\boldsymbol{i} \mathrel{\widehat{=}} \boldsymbol{i}_{k-1} \boldsymbol{i}_{k-2} \cdots \boldsymbol{i}_1 \boldsymbol{i}_0$, i.e., $\boldsymbol{i} = \sum_{j=0}^{k-1} Q^j(\boldsymbol{i}_j)$. We now show that the Q-adic expansion of $A^{(k)}(\boldsymbol{i})$ is given by $A^{(k)}(\boldsymbol{i}) \mathrel{\widehat{=}} A^{(1)}(\boldsymbol{i}_{k-1}) A^{(1)}(\boldsymbol{i}_{k-2}) \cdots A^{(1)}(\boldsymbol{i}_0)$. Plugging in the expansion of

$\boldsymbol{i}$ into Eq. (5.10), one gets $A^{(k)}(\boldsymbol{i}) = \Bigl(\sum_{j=0}^{k-1} AQ^j(\boldsymbol{i}_j)\Bigr) - A\boldsymbol{x}_k + \boldsymbol{x}_k$. On the other hand, one also has

$$\begin{aligned}
\sum_{j=0}^{k-1} Q^j(A^{(1)}(\boldsymbol{i}_j)) &= \sum_{j=0}^{k-1} Q^j\bigl(A(\boldsymbol{i} - Q\boldsymbol{v} + \boldsymbol{v}) + Q\boldsymbol{v} - \boldsymbol{v}\bigr) \\
&= \sum_{j=0}^{k-1} Q^j A(\boldsymbol{i}_j) + \sum_{j=0}^{k-1} \bigl(-AQ^{j+1}\boldsymbol{v} + AQ^j\boldsymbol{v}\bigr) + \sum_{j=0}^{k-1} \bigl(Q^j\boldsymbol{v} - Q^j\boldsymbol{v}\bigr) \\
&= \sum_{j=0}^{k-1} AQ^j(\boldsymbol{i}_j) \underbrace{-AQ^k\boldsymbol{v} + A\boldsymbol{v}}_{-A\boldsymbol{x}_k} + \underbrace{Q^k\boldsymbol{v} - \boldsymbol{v}}_{\boldsymbol{x}_k} \\
&= A^{(k)}(\boldsymbol{i}),
\end{aligned}$$

where the penultimate equality follows from $[A,Q]=0$ and the evaluation of the two telescoping sums. As in Theorem 5.13, one then obtains

$$\pi^{-1} \circ \theta_{\boldsymbol{i}} \circ \pi = \pi^{-1} \circ \theta_{\boldsymbol{i}_{k-1}} \circ \theta_{\boldsymbol{i}_{k-2}} \circ \cdots \theta_{\boldsymbol{i}_0} \circ \pi = \theta_{A^{(k)}(\boldsymbol{i})},$$

whenever $\boldsymbol{i} \mathrel{\widehat{=}} \boldsymbol{i}_{k-1}\boldsymbol{i}_{k-2}\cdots\boldsymbol{i}_0$ and $\pi^{-1} \circ \theta_{\boldsymbol{i}_s} \circ \pi = \theta_{A^{(1)}(\boldsymbol{i}_s)}$ for all $\boldsymbol{i}_s \in R$, which finishes the proof.

Remark The conditions $[A,Q]=0$ and $|A|\boldsymbol{x}_1 = \boldsymbol{x}_1$ in Theorem 5.19 are automatically satisfied if θ is a cubic substitution, i.e., $L_i = L$ for all $1 \leq i \leq d$, which means one can use Eq. (5.11) to check whether a given letter-exchange map works for any $A \in W_d$. For general θ, these relations are only satisfied for certain $A \in W_d$, e.g. reflections along coordinate axes, which means one needs a different tool to ascertain whether it is possible for other rigid motions to generate extended symmetries. For example, one can use Theorem 5.18 to exclude some symmetries.

Before we proceed, we need a higher-dimensional generalisation of Proposition 5.4 regarding aperiodicity. For this, we use the following result, which is formulated in terms of Delone sets. Here, S^{d-1} is the unit sphere in $\mathbb{R}^d$.

Theorem 5.20 ([6, Thm. 5.1]) *Let $\mathbb{X}(\varLambda)$ be the continuous hull of a repetitive Delone set $\varLambda \subset \mathbb{R}^d$. Let $\bigl\{\boldsymbol{b}_i \in S^{d-1} \mid 1 \leq i \leq d\bigr\}$ be a basis of $\mathbb{R}^d$ such that for each i, there are two distinct elements of $\mathbb{X}(\varLambda)$ which agree on the half-space $\{\boldsymbol{x} \mid \langle \boldsymbol{b}_i | \boldsymbol{x} \rangle > \alpha_i\}$ for some $\alpha_i \in \mathbb{R}^d$. Then one has that $\mathbb{X}(\varLambda)$ is aperiodic.* □

The proof of the previous theorem relies on the generalisation of the notion of proximality for tilings and Delone sets in $\mathbb{R}^d$, which is proximality along $\boldsymbol{s} \in S^{d-1}$; see [6, Sec. 5.5] for further details. Note that from a $\mathbb{Z}^d$-tiling generated by a rectangular substitution, one can derive a (coloured) Delone set $\varLambda$ by choosing a consistent control point for each cube (usually one of the corners or the centre). Primitivity guarantees that $\varLambda$ is repetitive and the notion of proximality extends trivially to coloured Delone sets using the same metric. The two hulls $\mathbb{X}(\varLambda)$ and X_θ are then mutually locally derivable, and the aperiodicity of one implies that of the other. We then have a sufficient criterion for the aperiodicity of X_θ in higher dimensions.

Proposition 5.21 *Let $\theta\colon \mathcal{A} \to \mathcal{A}^R$ be a d-dimensional rectangular substitution which is bijective and primitive. If there exist two legal blocks $u, v \in \mathcal{L}(\mathsf{X}_\theta)$ of side-length 2 in each direction such that u and v disagrees at exactly one position and coincides at all other positions, then the hull X_θ is aperiodic.*

Proof. The proof proceeds in analogy to Proposition 5.4. Here we choose the appropriate power to be $k = \operatorname{lcm}\left\{|\theta_{\boldsymbol{r}}|\colon \boldsymbol{r} = \sum_{i=1}^{d} r_i \boldsymbol{e}_i, r_i \in \{0, L_i - 1\}\right\}$. If we then place u and v at the origin, the resulting fixed points $x = \theta^\infty(u)$ and $x' = \theta^\infty(v)$ which cover $\mathbb{Z}^d$ will coincide at every sector except at the one where $u_{\boldsymbol{j}} \neq v_{\boldsymbol{j}}$. One can then choose $\boldsymbol{b}_i = \boldsymbol{e}_i$ and $\alpha_i = 0$ in Theorem 5.20, and for each i, x and x' to be the two elements which agree on a half-space, which guarantees the aperiodicity of X_θ. More concretely, x and x' are asymptotic, and hence proximal, along $\boldsymbol{e}_i$ for all $1 \leq i \leq d$.

Remark Obviously, one can have a lattice of periods of rank less than d in higher dimensions. An example would be when $\theta = \theta_1 \times \theta_2$, where θ_1 is the trivial substitution $a \mapsto aa, b \mapsto bb$ and θ_2 is Thue–Morse. Although θ_1 is itself not primitive, the product θ is and admits the legal blocks given in Figure 5.4, which generate fixed points that are $\mathbb{Z}\boldsymbol{e}_1$-periodic. If one requires that the shift component in $\operatorname{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ is $\mathbb{Z}^d$, one needs all elements of X_θ to be aperiodic in all cardinal directions, hence the stronger criterion in Proposition 5.21.

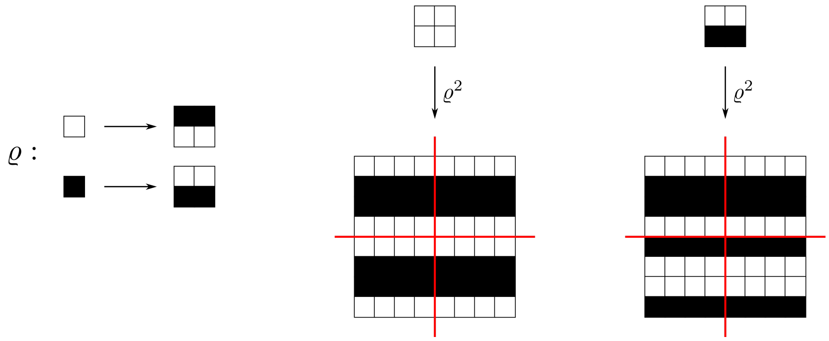


Figure 5.4: The image of two distinct blocks under θ coincide in the upper half-plane and are distinct in the lower half-plane. In the limit, these legal seeds generate two fixed points which are neither left nor right asymptotic with respect to $\sigma_{\boldsymbol{e}_1}$.

The next result is the analogue of Theorem 5.8 for extended symmetries, which holds in any dimension.

Theorem 5.22 *Given a finite group G and a subgroup P of the d-dimensional hyperoctahedral group W_d, there is an aperiodic, primitive, bijective d-dimensional substitution θ whose shift space satisfies*

$$\operatorname{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d) \cong \mathbb{Z}^d \times G$$
$$\operatorname{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d) \cong (\mathbb{Z}^d \rtimes P) \times G.$$

Proof. We start by taking a cursory look at the proof of Theorem 3.6 in [33]. For a given finite group G, we choose a generating set $S = \{s_1, \ldots, s_r\}$ that does not contain the identity, and build a substitution whose columns correspond to the left multiplication maps $L_{s_j}(g) = s_j \cdot g$, seen as permutations of the alphabet $\mathcal{A} = G$. These permutations generate the left Cayley embedding of G in the symmetric group on $|G|$ elements, whose corresponding centralizer, which induces all of the letter exchanges in $\operatorname{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$, is the right Cayley embedding of G generated by the maps $R_{s_j}(g) = g \cdot s_j$.

In what follows, we shall assume first that the group G is non-trivial, as the case in which G is trivial requires a slightly different construction. We also assume that the rectangular substitution we will construct engenders an aperiodic subshift, so that the group generated by the shifts is isomorphic to $\mathbb{Z}^d$. We delay the proof of this until later on, to avoid cluttering our construction with extraneous details.

Since $\operatorname{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ depends only on the columns of the underlying substitution and not their relative position, we shall construct a d-dimensional rectangular substitution θ with cubic support whose columns correspond to copies of the aforementioned L_{s_j}, placed in adequate positions along the cube. We start with a cube $R = [0, 2|S| + 2d + 1]^d$ of side length $2|S| + 2d + 2$, where the additional layer corresponding to the term 2 will be used below to ensure aperiodicity. This cube is comprised of $N = |S| + d + 1$ "shells" or "layers", which are the boundaries of the inner cubes $[j, 2|S| + 2d + 2 - j]^d$; we shall denote each of them by Λ_j, where j can vary from 0 to $N - 1$.

Fill the i-th inner shell Λ_{N-i} with copies of the column L_{s_i}, for all $1 \leq i \leq r$. This ensures that, as long as every other column is a copy of L_{s_j} for some j or an identity column, the automorphism group $\operatorname{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ of the corresponding subshift will be isomorphic to G, because in our construction the 2^d corners of the point will always be identity columns.

Now, note that N is chosen large enough so that the point $\boldsymbol{p} = (0, 1, \dots, d-1)$ lies in the outer $N - r \geq d$ shells and, moreover, the cube $[0, d-1]^d$ is contained in these outer shells as well. Thus, any permutation of the coordinates maps the cube $[0, d-1]^d$ to itself and, in particular, two different permutations map this point to two different points in this cube, that is, the orbit of $\boldsymbol{p}$ has $d!$ different points. Combining this with the fact that the mirroring maps send this cube to one of 2^d disjoint cubes (translations of $[0, d-1]^d$) in the corners of the larger cube $[0, N-1]^d$, it can be seen that W_d acts freely on the orbit of the point $\boldsymbol{p}$, that is, there is a bijection between the hyperoctahedral group W_d and the set $\operatorname{Orb}(\boldsymbol{p})$.

Next, choose a fixed $s_j \in S$ that is not the identity element of G, so that L_{s_j} is not an identity column. As P is a subgroup of W_d, it is bijectively mapped to the set $P \cdot \boldsymbol{p} = \{g \cdot \boldsymbol{p} : g \in P\}$. Place a copy of L_{s_j} in each position from $P \cdot \boldsymbol{p}$, and an identity column in every other position from $\operatorname{Orb}(\boldsymbol{p})$. Fill every remaining position in the cube with identity columns. This ensures that the group of letter exchanges will remain isomorphic to G, and, for each matrix $A \in W_d$ associated with some element $g \in P$, the map f_A given by the relation $f_A(x)_{\boldsymbol{n}} = x_{A\boldsymbol{n}}$ will be a valid extended symmetry, as a consequence of Theorem 5.19.

Since every other extended symmetry is a product of such an f_A with some letter-exchange map that has to satisfy the conditions given by Eq. (5.11) due to our construction, and L_{s_j} cannot be conjugate to the identity column, the only other extended symmetries are compositions of the already extant f_A with elements from $\operatorname{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$, i.e. $\operatorname{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)/\operatorname{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ has the equivalence classes of each f_A as its only elements. As the set of all f_A is an isomorphic copy of P contained in $\operatorname{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)$, we conclude that $\operatorname{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)$ is isomorphic to the semi-direct product $\operatorname{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d) \rtimes P$. However, since every letter exchanges from G commutes with every f_A trivially, this semi-direct product may be written as $\operatorname{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d) \cong (\mathbb{Z}^d \rtimes P) \times G$, as desired.

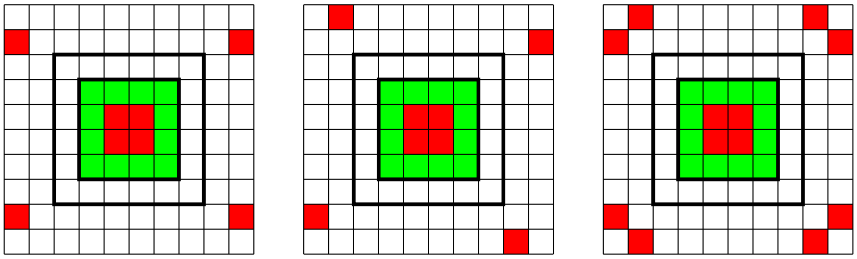


Figure 5.5: Examples of substitutions obtained by the above construction, for the Klein 4-group $C_2 \times C_2$, the cyclic group C_4 and the whole $W_2 = D_4$, respectively. The thicker lines mark the layer of identity columns separating the inner cube from the outer shell.

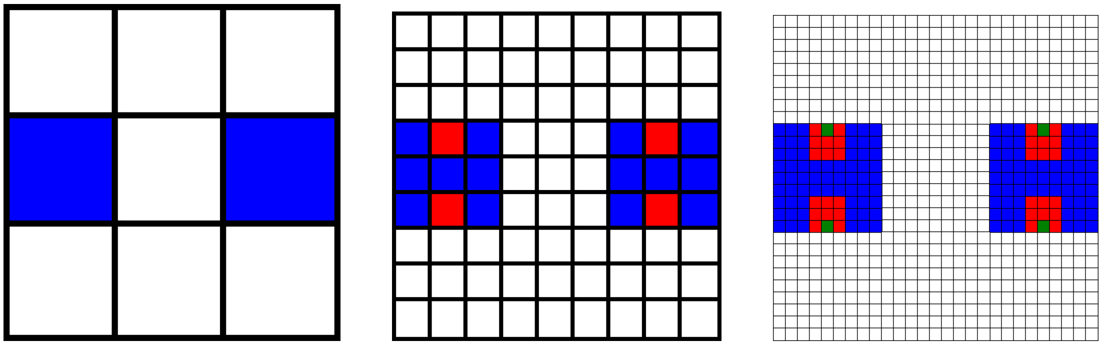
In the case where G is trivial, we may choose an alphabet with at least three symbols (to ensure that $S_{|\mathcal{A}|}$ is non-Abelian) and repeat the construction above with a collection of columns $\theta_0, \dots, \theta_{r-1}$ that generates some subgroup of $S_{|\mathcal{A}|}$ with trivial centralizer (e.g. the two generators of $S_{|\mathcal{A}|}$ itself). The rest of the proof proceeds in the same way.

To properly conclude the proof, we need to verify that the constructed substitution generates an aperiodic shift space. We focus on the case $d > 1$, as the one-dimensional case is a straightforward modification of the construction from Theorem 5.8. Since our d-dimensional cube has at least $d+1 \geq 3$ outer layers, we see that there is a $2 \times \cdots \times 2$ cube R_0 contained in the outer layers that does not overlap any of the 2^d cubes of size $d \times \cdots \times d$ on the corners nor the inner cube of size $2|S| \times \cdots \times 2|S|$. As a consequence, this cube R_0 contains only identity columns. Since we have a layer Λ_d consists only of identity columns directly enveloping the inner cube $\Lambda_{d+1} \cup \cdots \cup \Lambda_{d+|S|}$, the layer immediately following Λ_d is comprised only of non-identity columns, which are copies of the same bijection $\pi \colon \mathcal{A} \to \mathcal{A}$.. Thus, the 2^d corners of the hollow cube $\Lambda_d \cup \Lambda_{d+1}$ are $2 \times \cdots \times 2$ cubes $R_1, \cdots, R_{2^d}$ having exactly one non-identity column each, with this non-identity column τ being placed in every one of the 2^d possible positions on these cubes.

Since τ is not the identity, there must exist some $a \in \mathcal{A}$ such that $\tau(a) \neq a$. The previous discussion thus implies that there is an admissible pattern P_a of size $2 \times \cdots \times 2$ comprised only of copies of the symbol a, and $2^d + 1$ other admissible patterns $P_a^{(\boldsymbol{n})}$ that differ from P_a only in the position $\boldsymbol{n} \in [0,1]^d$. Using the proximality criterion from Proposition 5.21, we conclude that the subshift obtained is indeed aperiodic, as desired.

Remark An alternative Cantor-type construction, which produces the prescribed automorphism and extended symmetry groups, involves putting the non-trivial columns on the faces of R and labelling all columns in the interior to be the identity. Let G and P be given. From Theorem 5.8, there exists a substitution on $\mathcal{A}$ with $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d) = \mathbb{Z} \times G$. Let $\theta_0, \dots, \theta_{r-1}$ be the non-trivial columns of θ. Pick L to be large enough such that W_d acts freely on the faces of $R = [0, L-1]^d$. Choose $\boldsymbol{j}_0 \in R$ and consider the orbit of $\boldsymbol{j}_0$ under P, i.e., $\mathcal{O}_0 := P \cdot \boldsymbol{j}_0 = \{A \cdot \boldsymbol{j}_0 \mid A \in P\}$ where $A \cdot \boldsymbol{j} = A^{(1)}(\boldsymbol{j})$ as in Eq. (5.10) . Label all the columns in $\mathcal{O}_0$ with θ_0. We then expand R via $Q = \mathrm{diag}(L, \dots, L)$ to get the d-dimensional cube $Q(R)$ of side length L^2. Consider $\mathcal{B}_1 := Q(\mathcal{O}_0) + R$, pick $\boldsymbol{j}_1 \in \mathcal{B}_1$ and let $\mathcal{O}_1 = P \cdot \boldsymbol{j}_1$. Relabel all columns in $\mathcal{B}_1 \setminus \mathcal{O}_1$ with θ_0 and all columns in $\mathcal{O}_2$ with θ_1. One can continue this process until all needed column labels appear; see Figure 5.6 for a two-dimensional example.

Note that one has $\theta_{\boldsymbol{i}} = \theta_{A^{(1)}(\boldsymbol{i})}$ for all $A \in P$ and $\boldsymbol{i} \in R = [0, L-1]^d$ by construction, which



(a) $\mathcal{O}_0$ in blue (b) $\mathcal{B}_1 \setminus \mathcal{O}_1$ in blue, $\mathcal{O}_1$ in red (c) $\mathcal{B}_2 \setminus \mathcal{O}_2$ in green

Figure 5.6: An example in $\mathbb{Z}^2$ with three non-trivial columns θ_0 (blue), θ_1 (red) and θ_2 (green). Here, one has $G = \mathrm{cent}_{S_{|\mathcal{A}|}} \langle \theta_0, \theta_1, \theta_2 \rangle$ and $P \cong V_4$, where $V_4 \leq D_4 = W_2$ is the Klein-4 group.

means $\pi = \mathrm{id}$ gives rise to an element of $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)$ for all $A \in P$ by Theorem 5.19. No other extended symmetries can occur because all the location sets $\mathcal{B}_i$ only contain non-trivial labels and are P-invariant, whereas if $A \notin P$ induces an extended symmetry, one must have $\theta_{\boldsymbol{\ell}} = \mathrm{id}$ for some $\boldsymbol{\ell} \in \mathcal{B}_r$.

The resulting block substitution is primitive, since reordering the columns does not affect primitivity. It is also aperiodic because one has enough identity columns, and hence one can find the legal words required in Proposition 5.21. For example, in the constructed substitution in Figure 5.6, the legal seeds can be derived from the 2×2 block consisting of all identity columns (i.e. all white squares), and another one with all columns being the identity except at exactly one corner, where it is blue. This completes the picture and one has $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d) \cong \mathbb{Z}^d \times G$ and $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d) \cong (\mathbb{Z}^d \rtimes P) \times G$.

We now turn our attention to examples where the letter-exchange map π that generates $f \in \mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)$ is not given by the identity. In particular, in these examples, π does not commute with the letter-exchanges which correspond to the automorphisms in $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$. To avoid confusion, we will use letters for our substitution and the action of the hyperoctahedral group will be given by numbers, seen as permutations of the coordinates. Mirroring along a hyperplane will be denoted by m_i, where i is the respective coordinate.

Example We explicitly give a substitution whose automorphism group is $\mathrm{Aut}(\mathsf{X}_\varepsilon, \mathbb{Z}^d) = \mathbb{Z}^d \times C_3$ and build another C_3 component in $\mathrm{Sym}(\mathsf{X}_\varepsilon, \mathbb{Z}^d)$, which produces reversors of order 9. With the requirement on $\mathrm{Sym}(\mathsf{X}_\varepsilon, \mathbb{Z}^d)/\,\mathrm{Aut}(\mathsf{X}_\varepsilon, \mathbb{Z}^d)$, the space has to be at least of dimension $d \geq 3$.

$$\begin{aligned} \varepsilon_0 &= (a\,d\,g)(b\,e\,h)(c\,f\,i) & \varepsilon_2 &= (a\,b\,c)(d\,e\,f)(g\,h\,i) & \varepsilon_5 &= \mathrm{id} \\ \varepsilon_1 &= (a\,g\,d)(b\,h\,e)(c\,i\,f) & \varepsilon_3 &= (b\,c\,d)(e\,f\,g)(h\,i\,a) & & \\ & & \varepsilon_4 &= (c\,d\,e)(f\,g\,h)(i\,a\,b) & & \end{aligned}$$

Here one has $\mathrm{Aut}(\mathsf{X}_\varepsilon, \mathbb{Z}^3) = \mathbb{Z}^3 \times C_3$, which is generated by $(a\,d\,g)(b\,e\,h)(c\,f\,i)$. Depending on the positioning of the columns, $\mathrm{Sym}(\mathsf{X}_\varepsilon, \mathbb{Z}^3)$ can either be $\mathbb{Z}^3 \rtimes C_9$, $\mathbb{Z}^3 \rtimes C_3 \times C_3$ or $\mathbb{Z}^3 \times C_3$. The group $\mathbb{Z}^3 \rtimes C_3 \times C_3$ can be realised using the construction from Theorem 5.22. On the other hand, $\mathbb{Z}^3 \times C_3$ is obtained if one orbit of maximal size is labelled with just one non-identity ε_i once, and the rest with ε_0.

Note that $\pi = (a\,b\,c\,d\,e\,f\,g\,h\,i)$ sends $\varepsilon_2 \to \varepsilon_3 \to \varepsilon_4 \to \varepsilon_2$ and $\varepsilon_0 \to \varepsilon_0$, $\varepsilon_1 \to \varepsilon_1$. Taking the cube of $(a\,b\,c\,d\,e\,f\,g\,h\,i)$ gives $(a\,d\,g)(b\,e\,h)(c\,f\,i) \in \mathrm{cent}_{S_9}(G^{(1)})$, where $G^{(1)}$ is the group generated by the columns. This is consistent with the bounds calculated in Proposition 5.17. We will illustrate the positioning of a few elements following the construction in Theorem 5.22. We look at a position that has the maximum orbit size under W_3, for example $(0, 1, 2) \in R$. The orbit under C_3 is $(0, 1, 2), (1, 2, 0), (2, 0, 1)$, which is obtained by cyclically permuting the coordinates. We place ε_2 at position (0,1,2), ε_3 at position (1,2,0) and ε_4 at position (2,0,1). Since $\varepsilon_0, \varepsilon_1 \in \mathrm{cent}_{S_9}(G)$, we will position them each along a different orbit. All remaining positions will be filled with the identity to ensure that we cannot have additional automorphisms. We use Proposition 5.21 to ensure aperiodicity. It is easy to see that one gets the required patches by choosing the $2 \times 2 \times 2$ cube in the upper right corner from the first and second slices and the other one from the second and third. For this configuration, one has $\mathrm{Sym}(\mathsf{X}_\varepsilon, \mathbb{Z}^3) = \mathbb{Z}^3 \rtimes C_9$.

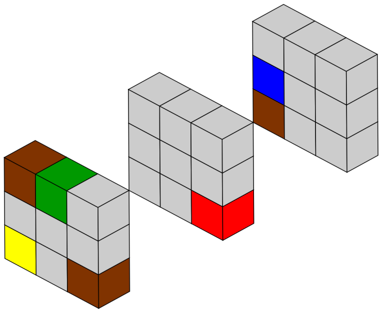


Figure 5.7: The gray cubes are filled with ε_5 (the identity). Yellow and brown can be filled by either $\varepsilon_0, \varepsilon_1$, respectively. Lastly, ε_2 is blue, ε_3 is green and ε_4 is red, where one has the obvious freedom in choosing the colours due to the C_3-symmetry.

Remark As a generalisation of Example 5.3.2, for any given cyclic groups C_n and C_k, we can construct a substitution θ in $\mathbb{Z}^n$, such that X_θ has the automorphism group $\mathbb{Z}^n \times C_k$ and its extended symmetry group is given by $(\mathbb{Z}^n \times C_k) \rtimes C_n$. More precisely, since the extended symmetry group contains an element of order nk, $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^n) = \mathbb{Z}^n \rtimes C_{nk}$. The substitution can be realised by the following columns

$$\begin{aligned}
\varepsilon_0 &= (a_1\, a_{k+1}\, \cdots\, a_{(n-1)k+1}) \cdots (a_k\, \cdots\, a_{nk}) \\
\varepsilon_i &= (a_i\, a_{i+1}\, \cdots\, a_{k-1+i})(a_{k+i}\, a_{k+i+1}\, \cdots\, a_{2k+i-1}) \cdots (a_{(n-1)k+i}\, a_{(n-1)k+i+1}\, \cdots\, a_{nk-1+i}) \\
\varepsilon_{n+1} &= \mathrm{id}
\end{aligned}$$

where i runs from 1 to n, where the values are seen modulo nk.

From the columns ε_i with $i \neq 0$ we can see that the centralizer can only be the permutation of the cycles limiting the centralizer to S_k, while ε_0 limits it further to be C_k, since this copy of S_k operates on the cycles independently and the centralizer of a cycle is just the cycle itself. The extended symmetry is realised by the permutation $(a_1 \cdots a_{kn})$ which maps ε_i to ε_{i+1}. Its orbit is determined by the action of $C_n \leq W_n$ on the positioning of the columns.

In the next example we illustrate how important it is to choose compatible structures for the letter-exchange map and the corresponding action in W_d.

Example We look at a four-letter alphabet with the following columns in Eq. (5.12) which generate S_4 as a subgroup of S_4, thus implying that the shift space to have a trivial centralizer. We plan to have $S_3 \cong \operatorname{Sym}(\mathsf{X}_\varepsilon, \mathbb{Z}^d)/\operatorname{Aut}(\mathsf{X}_\varepsilon, \mathbb{Z}^d)$, so we place the columns in a three-dimensional cube.

$$\begin{aligned}
&\varepsilon_0 = \mathrm{id} \qquad && \varepsilon_1 = (a\,b\,c\,d) \qquad && \varepsilon_4 = (a\,c\,d\,b) \\
& && \varepsilon_2 = (a\,b\,d\,c) && \varepsilon_5 = (a\,d\,b\,c) \\
& && \varepsilon_3 = (a\,c\,b\,d) && \varepsilon_6 = (a\,d\,c\,b)
\end{aligned} \tag{5.12}$$

The automorphism group is trivial since the columns generate S_4. Conjugation with $\tau = (c\,d)$ maps ε_1 to ε_2, just as any $\tau\kappa$, with $\kappa \in \operatorname{cent}_{S_4}(\varepsilon_2)$.

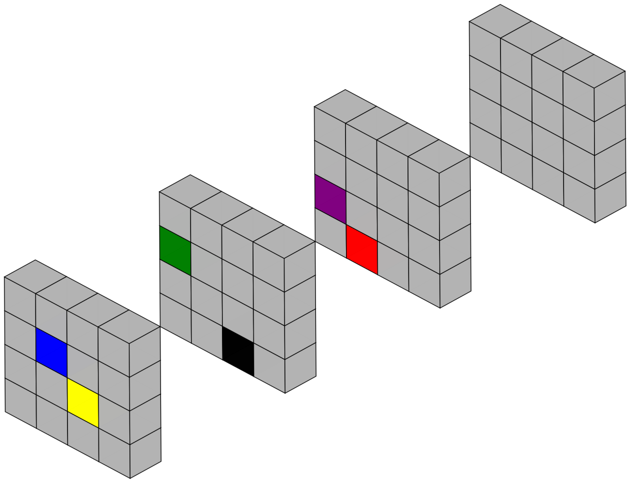


Figure 5.8: The columns assigned to the colors are as follows: ε_1 (blue), ε_2 (yellow), ε_3 (green) ε_4 (purple), ε_5 (black) and ε_6 (red).

Here $C_3 \rtimes C_2 \cong S_3$ is realised by $(b\,c\,d)(0\,1\,2)$ and $(c\,d)(0\,1)$. The transposition $(c\,d)$ cannot be realised in W_d by mirroring along an axis in the cube since that is not consistent with the interaction between $(b\,c\,d)$ and $(c\,d)$. This can be easily be seen by looking at mirroring along all hyperplanes.

We see that the diagram does not commute, thus there is no way to assign a single column to the vertex $(3, 1, 2)$. One can do this for all axes, which rules out the C_2^3 component in W_3, thus yielding $\operatorname{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d) = \mathbb{Z}^d \rtimes S_3$.

Remark One can also ask whether, starting with a group G, one can build the centralizer $\operatorname{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ and normalizer $\operatorname{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)$ organically from G, under a suitable embedding of G. Consider the Cayley embedding $G \hookrightarrow S_{|G|}$ as in Example 5.2.2. We know that $\operatorname{cent}_{S_{|G|}}(G) \cong G$ and $\operatorname{norm}_{S_{|G|}}(G) \cong G \rtimes \operatorname{Aut}(G)$; see [92]. Since the automorphisms of G are given by conjugation in $S_{|G|}$, they define letter-exchange maps which are compatible with reversors in $\operatorname{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)$. By choosing the dimension appropriately, one can construct a substitution θ on $\mathcal{A} = G$ such that the extended symmetry group is given by

$$\operatorname{Sym}(\mathsf{X}_\theta, \mathbb{Z}^{d(G)}) = \big(\mathbb{Z}^{d(G)} \times G\big) \rtimes \operatorname{Aut}(G),$$

where we choose $d(G)$ such that $\operatorname{Aut}(G) \leq W_{d(G)}$. This can always be done for $d(G) = |G|$, but depending on $\operatorname{Aut}(G)$, a smaller dimension is possible. Let $\pi \in \operatorname{Aut}(G)$ and let $A_\pi \in W_d$. The construction from the proof of Theorem 5.22 can be applied. Here, the orbits of A_π will not be filled with the same element, but with columns that are determined by π, i.e., $\theta_{A_\pi(\boldsymbol{i})} = \pi \circ \theta_{\boldsymbol{i}} \circ \pi^{-1}$, where π is seen as an element of $S_{|G|}$.

These series of examples with more complicated structure can be generalised for arbitrary groups G and P. Here, we have the following version of Theorem 5.22 where the letter exchange map is no longer $\pi = \mathrm{id}$, which we build from a specific set of columns.

Theorem 5.23 *Let H, P be arbitrary finite groups. Then for all $\ell \geq c(P)$, where $c(P)$ is a constant which depends only on the group P, there is a shift space X_θ originating from an aperiodic, primitive and bijective substitution θ such that*

$$\operatorname{Aut}(\mathsf{X}_\theta, \mathbb{Z}^\ell) = \mathbb{Z}^\ell \times H$$
$$\operatorname{Sym}(\mathsf{X}_\theta, \mathbb{Z}^\ell) = (\mathbb{Z}^\ell \times H) \rtimes P.$$

Proof. The proof will be divided into two parts, beginning with a manual for the construction of the substitution and a second part where we verify the claims made in the construction and check if the subshift has the desired properties.

- We first turn our attention to the construction of P which later is supposed to be isomorphic to $\operatorname{Sym}(\mathsf{X}_\theta, \mathbb{Z}^\ell)/\operatorname{Aut}(\mathsf{X}_\theta, \mathbb{Z}^\ell)$. For that purpose we embed $P \hookrightarrow S_\ell$ which is certainly possible for some ℓ. It is clear that there is a minimal $c(P) \in \mathbb{N}$ for which this embedding is possible, and that every $\ell \geq c(P)$ gives a valid embedding as well. This means the choice of ℓ has a lower bound, but can be increased arbitrarily. This chosen ℓ determines the dimension of the space $\mathbb{Z}^\ell$ where the subshift is constructed. Let us now fix a suitable ℓ, excluding $\ell = 2, 3, 6$ since we use want to use $\operatorname{Aut}_{S_\ell}(S_\ell) = \operatorname{Inn}_{S_\ell}(S_\ell) \cong S_\ell$ which does not hold for these values of ℓ; see [91].
- Next, we look for suitable columns for our substitution. Choose the set $T = \{\varepsilon_1, \cdots \varepsilon_k\}$ of all transpositions in S_ℓ, together with the identity column as the set of columns. T generates S_ℓ and the action of S_ℓ (viewed as the automorphism group) acts faithfully on T. From this, we get that $P \subset S_\ell \cong \operatorname{Inn}_{S_\ell}(S_\ell) \subset \operatorname{norm}_{S_\ell}(\{\varepsilon_1, \cdots, \varepsilon_k\})$. This is enough for now, since $P \subset \operatorname{norm}_{S_\ell}(\{\varepsilon_1, \cdots, \varepsilon_k\})$ and we can exclude the surplus later.
- Now, we compute the centralizer of the column group. In our current construction the centralizer is trivial, which is why we need to modify our columns. We do this by extending our alphabet $\{a, \cdots, \ell\}$ to $\{a_1, \cdots, a_{|H|}, b_1, \cdots, b_{|H|}, \cdots, \ell_1, \cdots, \ell_{|H|}\}$. We simply duplicate the cycles in each column: The permutations of the columns are mapped by $\rho \to \rho'$ sending $\varepsilon_i = (x\, y) \mapsto \varepsilon_i = (x_1\, y_1) \cdots (x_{|H|}\, y_{|H|})$.
- We embed $G \hookrightarrow S_{|H|}$ with the usual Cayley embedding. This group is only acting on the indices of the letters in the new alphabet. The action on the indices is applied to every $\{a, \ldots, \ell\}$, giving the final set of columns $\{\eta_1, \ldots, \eta_m\}$ added to the substitution ρ' giving a new substitution θ.
- The Cayley embedding guarantees that $\operatorname{cent}_{S_{|H|}}(G_\theta) \cong H$, where G_θ is the column group of θ. We can decrease the size of $\operatorname{Sym}(\mathsf{X}_\theta, \mathbb{Z}^\ell)/\operatorname{Aut}(\mathsf{X}_\theta, \mathbb{Z}^\ell)$ with the same arguments as in Theorem 5.22. This way we achieve a group $\operatorname{Sym}(\mathsf{X}_\theta, \mathbb{Z}^\ell)/\operatorname{Aut}(\mathsf{X}_\theta, \mathbb{Z}^\ell) \cong P$ where the letter exchange component π of the extended symmetries are not in $\operatorname{cent}_{S_{|H|}}(G_\theta)$.

Aperiodicity of X_θ can be easily obtained via proximal pairs. Regarding primitivity, it is sufficient to check the transitivity of G_θ and use Proposition 5.2. For any pair (x_j, y_k) of letters with indices chosen from the alphabet we need to find a $g \in G_\theta$ such that $gx_j = y_k$. Note that the permutation $(x_1\, y_1) \cdots (x_j\, y_j) \cdots (x_{|H|}\, y_{|H|}) \in G_\theta$ and maps x_j to y_j. Now we need to map y_j to y_k, which is an action solely on the indices. The mapping on the indices can be realized by the right embedding copy of H in $S_{|H|}$ and thus by an element composed of the columns $\{\eta_1, \cdots, \eta_m\}$.

Let us prove that the centralizer is indeed isomorphic to G. The centralizer of $G_{\{\varepsilon_1, \cdots \varepsilon_k\}}$ can only contain elements that are pure index permutations, since those columns generate S_ℓ. Since the structure of the cycles in each column are independent of the index, any index permutation is an element of $\mathrm{cent}_{S_\ell}(G_{\{\varepsilon_1, \cdots \varepsilon_k\}}) = S_\ell$.

We continue by determining $\mathrm{cent}_{S_{\ell|H|}}(G_{\{\eta_1, \cdots, \eta_m\}}) \bigcap S_\ell$. The group S_ℓ are the pure index switches and since $\eta_1, \cdots, \eta_m$ are the columns generated by the Cayley embedding of H into S_ℓ their centralizer is isomorphic to H.

The following rule lifts an automorphism h' on G_ρ to h on G_θ .

$$h(\varepsilon_i) = h'(\varepsilon)_i$$

Thus $S_{|H|} \leq \mathrm{Aut}_{S_{\ell|H|}}(G_{\{\varepsilon_1, \cdots, \varepsilon_k\}})$. It is sufficient to prove that the automorphism group did not decrease in size by the addition of the columns $(\eta_1, \cdots, \eta_m)$. Then we can use the geometric placement of the columns in Theorem 5.22 in the substitution to exclude any unwanted W_d-component. Any lifted automorphism h still only maps the letters and fixes the indices. Since the cycles in any $\eta_1, \cdots, \eta_m$ contain only the same letter with different indices and the index structure is independent of the letter, every h is in $\mathrm{cent}_{S_{|H|\ell}}(G_{\{\eta_1, \cdots, \eta_m\}})$ and surely legal. Thus it is an automorphism on the whole of G_θ.

Remark Theorems 5.22 and 5.23 fall under realisation theorems for shift spaces. The most general current result along this vein known to the authors is that of Cortez and Petite, which states that every countable group G can be realised as a subgroup $G \leq \mathrm{Sym}(X, \mathit{\Gamma})$, where $\mathrm{Sym}(X, \mathit{\Gamma})$ is the normalizer of the action of a free abelian group $\mathit{\Gamma}$ on an aperiodic minimal Cantor space X; see [27].

5.4. Concluding remarks

While the higher-dimensional criteria in Theorems 5.19 and 5.18, which confirm or rule out the existence of extended symmetries, are rather general, it remains unclear how to find a way to extend this to a larger (possibly all) class of systems, with no constraints on the geometry of the supertiles. This is related to a question of determining whether, given a substitution in $\mathbb{Z}^d$ (or $\mathbb{R}^d$), one can come up with an algorithm which decides whether there is a simpler substitution which generates the same or a topologically conjugate hull, which is easier to investigate. This is exactly the case for the two-dimensional Thue–Morse substitution in Figure 5.2. Such an issue is non-trivial both in the tiling and the subshift context; see [26, 35, 52].

Note that the letter-exchange map $\pi \in S_{|\mathcal{A}|}$ in Theorem 5.19 always induces a conjugacy between columns whenever it generates a valid reversor. It would be interesting to know whether outer automorphisms in this case can yield valid reversors for a bijective substitution subshift in $\mathbb{Z}^d$, for example for those whose geometries are not covered by Theorems 5.19 and 5.18. For instance, $\mathrm{Aut}(S_6)$ contains elements which are not realised by conjugation.

Another natural question would be to determine other possibilities for $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ and $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)$ outside the class of bijective, constant-length substitutions. Here, the higher-dimensional generalisations of the Rudin–Shapiro substitution would be good candidates; see [40]. There are also substitutive planar tilings with $|\operatorname{Sym}(X, \mathbb{Z}^2)/\operatorname{Aut}(X, \mathbb{Z}^2)| = D_6$, which arises from the hexagonal symmetry satisfied by the underlying tiling. For these classes, and in the examples treated above, the simple geometry of the tiles introduces a form of rigidity which leads to $\mathrm{Sym}(X, \mathbb{Z}^2)$ being a finite extension of $\mathrm{Aut}(X, \mathbb{Z}^2)$; see [13, Sec. 5] for the notion of hypercubic shifts. There are substitution tilings whose expansive maps Q are no longer diagonal matrices, and whose supertiles have fractal boundaries; compare [42, Ex. 12], which allows more freedom in terms of admissible elements of $\mathrm{GL}(d, \mathbb{Z})$ which generate reversors. This raises the following question:

Question What is the weakest condition on the shift space/tiling dynamical system X which guarantees $\big[\mathrm{Sym}(X, \mathbb{Z}^d) : \mathrm{Aut}(X, \mathbb{Z}^d)\big] < \infty$?

This is always true in one dimension regardless of complexity, since either the subshift is reversible or not, but is non-trivial in higher dimensions because $|\mathrm{GL}(d, \mathbb{Z})| = \infty$ for $d > 1$, so infinite extensions are possible; see [4]. We suspect that this is connected to the notions of linear repetitivity, finite local complexity, and rotational complexity; compare [13, Cor. 4] and [52]. For inflation systems, the compatibility condition $[A, Q] = 0$ in Theorem 5.19 might also be necessary in general when the maximal equicontinuous factor (MEF) has an explicit form.

5.5. Acknowledgements

The authors would like to thank Michael Baake for fruitful discussions and for valuable comments on the manuscript. AB is grateful to ANID (formerly CONICYT) for the financial support received under the Doctoral Fellowship ANID-PFCHA/Doctorado Nacional/2017-21171061. NM would like to acknowledge the support of the German Research Foundation (DFG) through the CRC 1283.

Part III

Dynamical systems of algebraic origin

Chapter 6

Number-theoretical k-free shifts

This chapter is largely an adaptation of *Number-theoretic positive entropy shifts with small centraliser and large normaliser* [4], a joint work with Michael Baake, Christian Huck, Marius Lemańczyk and Andreas Nickel. Some commentary has been added regarding extensions of this work that appear in the (not yet published) joint work with Michael Baake and Andreas Nickel, *On the stabiliser of some number-theoretic shift spaces* [5].

6.1. Introduction

Shift spaces under the action of $\mathbb{Z}^d$ form a much-studied class of dynamical systems, both for $d = 1$, compare [71], and for $d \geq 2$. In the latter case, much less is known in terms of general classifications, and even subclasses such as those of algebraic origin [90] are still rather enigmatic, despite displaying fascinating facets that have been analyzed intensely. In particular, one is looking for interesting topological invariants to help analyse the jungle, and quite a bit of progress in this direction has been made recently.

Among the available tools are the automorphism group of a shift space and its various siblings and generalisations; see [3,27,29,30,33,36] and references therein. Here, we adopt the point of view of [3,13] to analyse both the (topological) centralizer (denoted by $\mathrm{Aut}(X, \mathbb{Z}^d)$ below) *and* the normalizer of the shift space, the latter denoted by $\mathrm{Sym}(X, \mathbb{Z}^d)$, as this pair can be quite revealing as soon as $d \geq 2$. In fact, both the topological setting and the extension to higher dimensions go beyond some of the initial studies [46,61] that specifically looked at reversibility in the measure-theoretic setting for $d = 1$; see [3, 82, 87] and references therein for more on the early reversibility results. Further, the groups $\mathrm{Aut}(X, \mathbb{Z}^d)$ and $\mathrm{Sym}(X, \mathbb{Z}^d)$ are often explicitly accessible, both for systems of low complexity, where $\mathrm{Aut}(X, \mathbb{Z}^d)$ is often minimal due to some form of topological rigidity, and beyond, where other rigidity mechanisms of a more algebraic nature emerge.

Below, we consider binary shift spaces of number-theoretic origin, as motivated by recent progress on $\mathcal{B}$-free systems and weak model sets; see [9, 36, 37, 56] and references therein. By way of characteristic examples with pure point spectrum, we demonstrate that positive topological entropy may very well be compatible with small or trivial centralizers, which means that $\mathrm{Aut}(X, \Gamma)$ agrees with the underlying lattice Γ (meaning a co-compact discrete subgroup of $\mathbb{R}^d$, which we often assume to be $\mathbb{Z}^d$) or a finite-index extension thereof, but also that such systems may have considerably larger normalizers, which is particularly interesting for $d \geq 2$. In fact, as shown in [13,22], it is the group $\mathrm{Sym}(X, \Gamma)$ that captures some obvious symmetries, as visible from the chair tiling and related shift spaces with their pertinent geometric symmetries. Also, the computability of $\mathrm{Aut}(X, \Gamma)$ and $\mathrm{Sym}(X, \Gamma)$ in these cases can be an advantage over some of the more general, abstract (semi-)groups that are presently attracting renewed attention.

The paper is organised as follows. After the introduction of some concepts and notions in Section 6.2, we set the scene with the well-known example of the visible lattice points of $\mathbb{Z}^2$ in Section 6.3, leading to Proposition 6.3 and Corollary 6.4, which in particular show that one has $\mathrm{Sym}(\mathsf{X}_V, \mathbb{Z}^2) = \mathbb{Z}^2 \rtimes \mathrm{GL}_2(\mathbb{Z})$. Then, Section 6.4 states and proves this for $\mathbb{Z}^d$ with $d \geq 2$ (Theorem 6.5) and introduces the general framework of lattice-based shift spaces, which can often be characterised by a rather powerful admissibility condition for its elements (Proposition 6.7). Then, under some mild assumptions, the normalizers are always maximal extensions of the corresponding centralizers (Theorem 6.8), with elements that are affine mappings (Corollary 6.9).

Section 6.5 explains the general number-theoretic setting of an algebraic $\mathcal{B}$-free system in higher dimensions, based on the classic Minkowski embedding of (commutative) maximal orders and their ideals as lattices in $\mathbb{R}^d$ for a suitable d. Here, Theorem 6.12 states the results on the triviality of $\mathrm{Aut}(X, \mathbb{Z}^d)$ and the direct product nature of $\mathrm{Sym}(X, \mathbb{Z}^d)$, which are true under a coprimality condition of the ideals chosen for $\mathcal{B}$ and a mild convergence condition, together known as the Erdős property, in generalisation of the one-dimensional notion [36] from $\mathcal{B}$-free integers.

Sections 6.6 and 6.7 then cover some paradigmatic examples from quadratic number fields. In the complex case, we treat the shift spaces generated by the k-free Gaussian or the k-free Eisenstein integers (Theorems 6.17 and 6.18). In both cases, $\mathrm{Sym}(X, \mathbb{Z}^2)$ is the extension of $\mathrm{Aut}(X, \mathbb{Z}^2) \cong \mathbb{Z}^2$ by a maximal finite subgroup of $\mathrm{GL}_2(\mathbb{Z})$, which is substantially different from the case of the visible lattice points. Finally, in the real case, we consider k-free integers in the maximal order of $\mathbb{Q}(\sqrt{m}\,)$ for $m \in \{2, 3, 5\}$. Here, Theorem 6.21 states that $\mathrm{Sym}(X, \mathbb{Z}^2)$ is the semi-direct product of $\mathrm{Aut}(X, \mathbb{Z}^2) \cong \mathbb{Z}^2$ with a non-trivial *infinite* subgroup of $\mathrm{GL}_2(\mathbb{Z})$, which can be given a clear interpretation in terms of algebraic number theory. The latter case, which is the first example of this type to the best of our knowledge, is intermediate between known examples from inflation tilings and shifts such as that generated by the visible lattice points. Thus, it looks particularly promising for future work and extensions to general number fields.

6.2. Preliminaries

Let $\Gamma \subset \mathbb{R}^d$ be a **lattice** in d-space, that is, a discrete and co-compact subgroup of $\mathbb{R}^d$. Below, we will be working with the full shift (or configuration) space $\{0, 1\}^{\Gamma}$, equipped with the standard product topology, and certain of its closed subspaces (called subshifts or simply shifts). We will generally use X to denote such a shift space, refering to either the full shift or the subshift under consideration. When the situation is independent of the geometry of the lattice, we will choose $\Gamma = \mathbb{Z}^d$ for simplicity. Any element $x \in X$ can also be viewed as a subset of Γ, by taking the support of x, that is, by mapping x to

$$U_x := \mathrm{supp}(x) = \{n \in \Gamma : x_n = 1\} \subseteq \Gamma.$$

Conversely, any point set $U \subseteq \Gamma$ can be viewed as a configuration, by mapping it to $x_U = 1_U$, that is, to its characteristic function. As usual, X admits a continuous action of Γ on it, defined by $T\colon \Gamma \times X \longrightarrow X$, where $T(t,x) = T_t(x)$ with $\bigl(T_t(x)\bigr)_n := x_{n+t}$. When working with $\mathbb{Z}^d$, we shall usually refer to its standard basis as $\{e_1, \ldots, e_d\}$, and align this with the elementary shift action of the d commuting shift operators T_{e_i}. For the action of $\mathbb{Z}^d$ in this case, with $t = (t_1, \ldots, t_d)$, this simply means $T_t(x) = T_{e_1}^{t_1} \cdots T_{e_d}^{t_d}(x)$ for all $x \in \{0,1\}^{\mathbb{Z}^d}$.

Likewise, there is an action of Γ on its subsets defined by $\alpha_t(U) = t + U := \{t + u : u \in U\}$. It is easy to check that $U_{T_t(x)} = \alpha_{-t}(U_x)$. If we denote the power set of Γ by Ω, we thus get the commutative diagram

$$\begin{tikzcd}
X \arrow[r, "T_t"] \arrow[d, "\gamma"'] & X \arrow[d, "\gamma"] \\
\Omega \arrow[r, "\alpha_{-t}"] & \Omega
\end{tikzcd} \tag{6.1}$$

where γ is the mapping defined by $x \mapsto U_x$. This is a homeomorphism if we equip Ω with the **local topology**, where two subsets of Γ are ε-close to one another when they agree on the ball of radius $1/\varepsilon$ around 0. Consequently, by slight abuse of notation, we will not distinguish these two points of view whenever the context is clear. This means that we will consider a subset $U \subseteq \Gamma$ simultaneously as a configuration, and vice versa.

In this spirit, we can also consider the group of lattice automorphisms, $\mathrm{Aut}(\Gamma) \cong \mathrm{GL}_d(\mathbb{Z})$. Indeed, if $\mathrm{Homeo}(X)$ denotes the group of homeomorphisms of X, any $M \in \mathrm{Aut}(\Gamma)$ induces an element $h_M \in \mathrm{Homeo}(X)$, where

$$\bigl(h_M(x)\bigr)_n := x_{M^{-1}n}. \tag{6.2}$$

In fact, the mapping $M \mapsto h_M$ defines an injective group homomorphism. Here, one can check that $U_{h_M(x)} = MU_x$, so the counterpart to (6.1) is the commutative diagram

$$\begin{tikzcd}
X \arrow[r, "h_M"] \arrow[d, "\gamma"'] & X \arrow[d, "\gamma"] \\
\Omega \arrow[r, "M"] & \Omega
\end{tikzcd} \tag{6.3}$$

which makes calculations with elements of the form h_M more convenient in the formulation with subsets. From now on, we identify X and Ω, and use the symbol X for both; when we intend to explicitly show the dependence between X and V, we use the symbol X_V (with a different typeface), to follow the previous notational conventions. To ease the understanding, we will normally use x, y for configurations and U, V for sets.

Remember that a **point set** $S \subset \mathbb{R}^d$, by which we mean an at most countable union of singleton sets, is said to have **natural density** δ if

$$\mathrm{dens}(S) = \lim_{r\to\infty} \frac{|S \cap B_r|}{\mathrm{vol}(B_r)}$$

exists (and equals δ), where B_r denotes the closed ball of radius r around 0. One can use other sets for averaging, as long as they are centered around 0 and satisfy some condition of Følner or van Hove type; see [6, 11] for details.

Below, we shall need the following simple result on sublattices of a given lattice, where the term **sublattice** is meant to include the property that the corresponding index is finite.

Fact 6.1 *Let Γ be a lattice in $\mathbb{R}^d$, and let Γ_1 and Γ_2 be sublattices of Γ, with corresponding indices m_1 and m_2, respectively. Then, $\Gamma_1 \cap \Gamma_2$ and $\Gamma_1 + \Gamma_2$ are sublattices of Γ as well.*

Further, if the indices m_1 and m_2 are coprime, one has $\Gamma_1 + \Gamma_2 = \Gamma$, which implies that Γ_1 meets all cosets of Γ_2 and vice versa.

Proof. If $[\Gamma : \Gamma_i] = m_i$, one has $m_i \Gamma \subseteq \Gamma_i$ by standard arguments, which implies

$$m_1 m_2 \Gamma \,\subseteq\, \Gamma_1 \cap \Gamma_2 \,\subseteq\, \Gamma_1 + \Gamma_2 \,\subseteq\, \Gamma .$$

The sublattice property for $\Gamma_1 \cap \Gamma_2$ and $\Gamma_1 + \Gamma_2$ is then clear.

The next statement is a consequence of what is sometimes referred to as the diamond isomorphism theorem, but can also be seen directly as follows. Set $n = [\Gamma : (\Gamma_1 + \Gamma_2)]$ and $n_i = [(\Gamma_1 + \Gamma_2) : \Gamma_i]$. Then, for $i \in \{1, 2\}$,

$$m_i \,=\, [\Gamma : \Gamma_i] \,=\, [\Gamma : (\Gamma_1 + \Gamma_2)][(\Gamma_1 + \Gamma_2) : \Gamma_i] \,=\, n n_i ,$$

which implies $n \mid \gcd(m_1, m_2) = 1$ and thus $\Gamma_1 + \Gamma_2 = \Gamma$. The final implication for the cosets is a now simple consequence.

As described in Chapter 3, one of the main tools in the description of shift spaces are **sliding block codes**, also known as **block maps**; see [71] for background. Remember that, given two subshifts $X \subseteq \mathcal{A}^{\mathbb{Z}^d}$ and $Y \subseteq \mathcal{B}^{\mathbb{Z}^d}$ over finite alphabets $\mathcal{A}$ and $\mathcal{B}$, a continuous mapping $h \colon X \longrightarrow Y$ is called a **block map** if there is a non-negative integer[1] ℓ such that, for every $x \in X$ and all $n \in \mathbb{Z}^d$, the image $y = h(x)$ at position n is fully determined from the patch $\big\{ x_{n+m} : m \in [-\ell, \ell]^d \big\}$, that is, from the knowledge of x within a d-cube of sidelength 2ℓ centered at n. In other words, the action of h can be seen as the result of a sliding block code $\phi = \phi_h$ that, for some fixed $\ell \in \mathbb{N}_0$, maps a cubic block of $(2\ell + 1)^d$ symbols from $\mathcal{A}$ to a single letter from $\mathcal{B}$, positioned at the center of the block (which can easily be modified when needed). This is the symbolic version of a **local derivation rule** from discrete geometry [6, Sec. 5.2]. An important result that we shall need repeatedly is the **Curtis–Hedlund–Lyndon theorem** (CHL): If a continuous mapping $h \colon X \longrightarrow Y$ intertwines the shift action on X and Y, it must be a block map based on some code ϕ of the above type [71, Thm. 6.2.9].

Below, as in the previous chapters, we shall only be interested in subshifts on which the action of Γ is **faithful**, which means that the subshift contains non-periodic elements. In this context, it is also natural to consider the affine lattice group $\Gamma \rtimes \mathrm{Aut}(\Gamma)$, whose elements (t, M) act on $\mathbb{R}^d$ via $(t, M)(y) := My + t$, and correspondingly on X. In this formulation, the group multiplication is $(t, M)(s, N) = (t + Ms, MN)$, with neutral element $(0, \mathbb{1})$ and inverse elements $(t, M)^{-1} = (-M^{-1} t, M^{-1})$. This group will become important later.

Further notation and concepts can now better be introduced along a paradigmatic example, which will simultaneously motivate the various extensions to follow.

6.3. Visible lattice points and their shift space

Consider the **visible points** V of the square lattice, $\mathbb{Z}^2$, which are defined as

$$V := \{(m, n) \in \mathbb{Z}^2 : \gcd(m, n) = 1\}.$$

[1] The least possible ℓ is the **radius**, as described in Chapter 3.

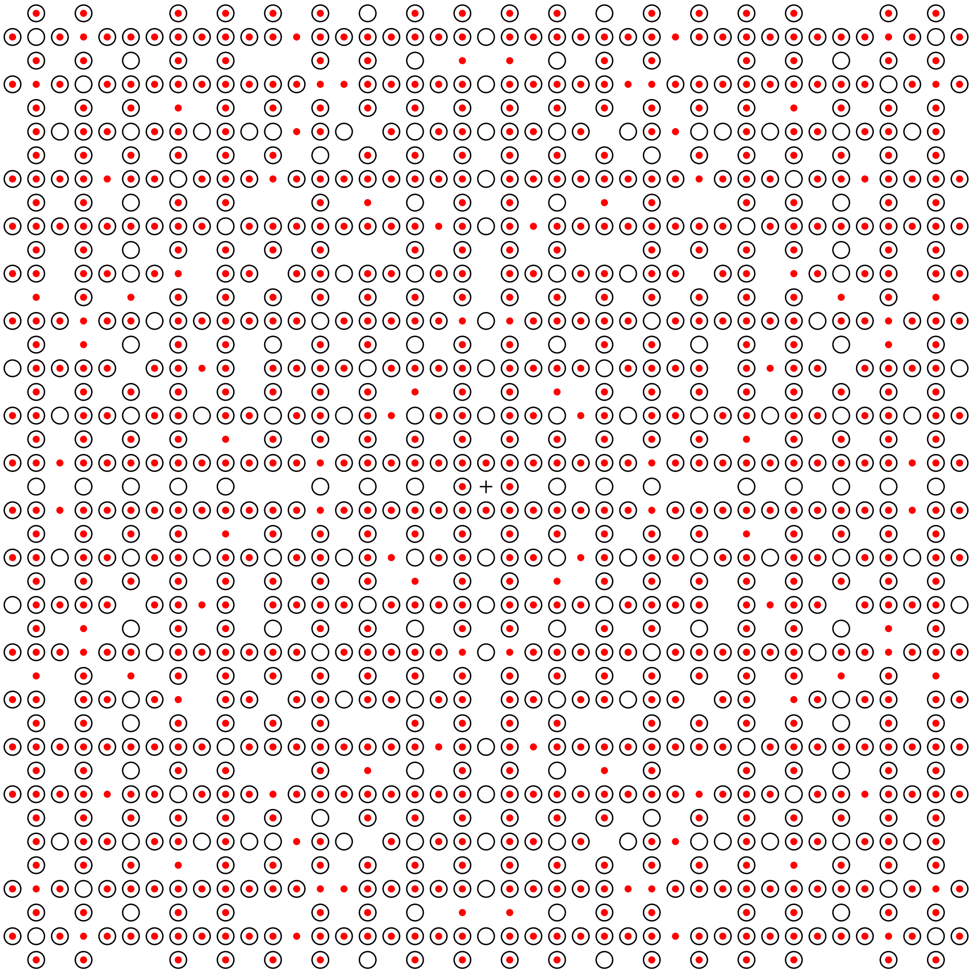


Figure 6.1: Central patch of the visible points of $\mathbb{Z}^2$ (dots) and of the square-free Gaussian integers (circles). The cross in the center marks the origin.

They are also known as the primitive points, and are used in many places; see also the cover page of [1]. Clearly, one has $V = \mathbb{Z}^2 \setminus \bigcup_{p\in\mathbb{P}}(p\mathbb{Z}^2)$, where $\mathbb{P}$ denotes the set of rational primes. Figure 6.1 shows a finite patch around the origin, in comparison with another set that will be discussed later, in Section 6.6. Let us recall some well-known properties of V; see [6, 11] and references therein for background and further results.

Fact 6.2 *The set V is uniformly discrete, but not relatively dense. In particular, V contains holes of unbounded inradius that repeat lattice-periodically. Yet, it satisfies $V - V = \mathbb{Z}^2$ and has natural density $\operatorname{dens}(V) = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}$, where $\zeta(s)$ is Riemann's zeta function.*

Furthermore, the set V is pure point diffractive, with the diffraction measure being invariant under the action of the affine group $\mathbb{Z}^2 \rtimes \mathrm{GL}_2(\mathbb{Z})$. □

Now, let $\mathsf{X}_V := \overline{\mathbb{Z}^2 + V}$ be the orbit closure of V under the shift (or translation) action of $\mathbb{Z}^2$, where the closure is taken in the standard product topology, also known as the **local topology** due to its geometric interpretation: Two configurations (or subsets) are close if they agree on a large neighborhood of $0 \in \mathbb{Z}^2$. In particular, since V has holes of arbitrary size, one immediately obtains that $\varnothing \in \mathsf{X}_V$, where $\varnothing$ is the empty set and represents the all-0 configuration. Clearly, X_V is a compact space, which is canonically identified with a subshift in $\{0,1\}^{\mathbb{Z}^2}$, and $\bigl(\mathsf{X}_V, \mathbb{Z}^2\bigr)$ is a topological dynamical system.

Call a subset of $\mathbb{Z}^2$ **admissible** if it misses at least one coset modulo $p\mathbb{Z}^2$ for any $p \in \mathbb{P}$. One easily verifies that the set of admissible sets constitutes a subshift of $\{0,1\}^{\mathbb{Z}^2}$ as well, denoted by $\mathbb{A}$. Since the set V by the remark above misses the zero coset modulo each $p\mathbb{Z}^2$, one readily verifies that the elements of X_V are admissible, so $\mathsf{X}_V \subseteq \mathbb{A}$. In fact, it was shown in [8, Lemma 4] that V shows all cosets except the zero coset modulo each $p\mathbb{Z}^2$, and is thus a maximal element of X_V. Further, one has the following result, the first part of which will be generalised below on the basis of Propositions 6.7 and 6.11.

Proposition 6.3 ([8]) *The space X_V coincides with the shift space of admissible sets, $\mathbb{A}$. In particular, X_V is hereditary (closed under the formation of subsets). The topological dynamical system $\bigl(\mathsf{X}_V, \mathbb{Z}^2\bigr)$ has topological entropy $\frac{6}{\pi^2}\log(2)$.*

With respect to the existing natural frequency measure ν_{M}, which is also known as the Mirsky measure, the measure-theoretic dynamical system $\bigl(\mathsf{X}_V, \mathbb{Z}^2, \nu_{\mathrm{M}}\bigr)$ has pure point dynamical spectrum, but trivial topological point spectrum.

The measure ν_{M} is ergodic for the $\mathbb{Z}^2$-action, and V is a generic element for ν_{M} in X_V. Moreover, the measure-theoretic entropy for ν_{M} vanishes. □

The characterisation of a number-theoretic shift space via an admissibility condition was originally observed by Sarnak for the square-free integers, and later extended to Erdős $\mathcal{B}$-free numbers in [37] and generalised to the lattice setting in [84]. Since this step is vital to us, we later present a streamlined version of the proof that covers the generality we need.

Remark The generating shifts induce unitary operators on the Hilbert space $L^2(\mathsf{X}_V, \nu_{\mathrm{M}})$, and the simultaneous eigenfunctions form a basis of this space [9, 10]. Except for the trivial one, no other eigenfunction is continuous. However, as follows from a recent result by Keller [59], see also the discussion in [9, 14], there is a subset of X_V of full measure on which the eigenfunctions *are* continuous. This is related to the fact that V is a weak model set of maximal density [9] in the cut and project scheme $(\mathbb{R}^2, H, \mathcal{L})$ with compact internal group $H = \prod_{p\in\mathbb{P}} \mathbb{Z}^2/p\mathbb{Z}^2$ and the lattice $\mathcal{L}$ being the diagonal embedding of $\mathbb{Z}^2$ into $\mathbb{R}^2 \times H$. It is an interesting open problem to understand the missing null set, and to connect it with the rather intricate relation between the topological and the measure-theoretic structure of this dynamical system.

Let $\mathrm{Homeo}(\mathsf{X}_V)$ be the group of all homeomorphisms of X_V, irrespective of whether they commute with the generators T_1, T_2 of the $\mathbb{Z}^2$-action or not. The translation action of $\mathbb{Z}^2$ on X_V is faithful, wherefore we have $\mathcal{G} := \langle T_1, T_2\rangle \cong \mathbb{Z}^2$. Clearly, X_V is not the full shift, and $\varnothing$ is the only fixed point of X_V under the translation action, since $\mathbb{Z}^2$ (as the all-1 configuration) is not an element of X_V.

The **automorphism group** of X_V, see [3] and references therein for background, is

$$\mathrm{Aut}(\mathsf{X}_V, \mathbb{Z}^2) := \mathrm{cent}_{\mathrm{Homeo}(\mathsf{X}_V)}(\mathcal{G}) = \{H \in \mathrm{Homeo}(\mathsf{X}_V) : GH = HG \text{ for all } G \in \mathcal{G}\},$$

which clearly contains $\mathcal{G}$ as a normal subgroup. This centralizer is sometimes called the **symmetry group** of the subshift, denoted as $\mathcal{S}(\mathsf{X}_V)$, most prominently in these works that follow the Smale convention, where the notation $\mathrm{Aut}(X)$ is reserved for what we denote as $\mathrm{Homeo}(X)$. For any $S \in \mathrm{Aut}(\mathsf{X}_V, \mathbb{Z}^2)$, $S(V)$ has a dense shift orbit in X_V (as V has dense orbit by definition). Moreover, one has $S(\varnothing) = \varnothing$ since $S(\varnothing)$ can also be seen as a fixed point under the translation action.

Now, consider an arbitrary $S \in \mathrm{Aut}(\mathsf{X}_V, \mathbb{Z}^d)$. By the CHL theorem, there is a block code (or map) $\phi\colon \{0,1\}^{[-\ell,\ell]^2} \longrightarrow \{0,1\}$ of a suitable size (parameterised by ℓ) such that, for any $x \in \mathsf{X}_V$, the value of Sx at a position $k \in \mathbb{Z}^2$ is given by the value under ϕ of the corresponding block of x around this very position, which we call its **center**. This means

$$(Sx)_k \, = \, \phi\bigl(x_{[k+[-\ell,\ell]^2]}\bigr),$$

where $x_{[k+[-\ell,\ell]^2]}(m) = x_{k+m}$ for $m \in [-\ell,\ell]^2$. Since $S(\varnothing) = \varnothing$, it is clear that $\phi\bigl(0_{[-\ell,\ell]^2}\bigr) = 0$.

Next, following [3, 13], we define the **extended symmetry group** as

$$\mathrm{Sym}(\mathsf{X}_V, \mathbb{Z}^2) := \mathrm{norm}_{\mathrm{Homeo}(\mathsf{X}_V)}(\mathcal{G}) \, = \, \{H \in \mathrm{Homeo}(\mathsf{X}_V) : H\mathcal{G}H^{-1} = \mathcal{G}\},$$

which contains both $\mathcal{G}$ and $\mathrm{Aut}(\mathsf{X}_V, \mathbb{Z}^2)$ as normal subgroups. Every $H \in \mathrm{Sym}(\mathsf{X}_V, \mathbb{Z}^d)$ must satisfy $H(\varnothing) = \varnothing$, as $H(\varnothing)$ can once again be shown to be fixed under any element of $\mathcal{G}$. Since every extended symmetry induces an automorphism of $\mathcal{G} \cong \mathbb{Z}^2$ via the conjugation action, $\mathrm{Sym}(\mathsf{X}_V, \mathbb{Z}^2)$ can at most be a group extension of $\mathrm{Aut}(\mathsf{X}_V, \mathbb{Z}^2)$ by $\mathrm{Aut}(\mathbb{Z}^2) = \mathrm{GL}_2(\mathbb{Z})$.

Let us state the final result for this specific example, which is a special case of our more general statement (Theorem 6.5) in the next section.

Corollary 6.4 *For the topological dynamical system $\bigl(\mathsf{X}_V, \mathbb{Z}^2\bigr)$, the automorphism group is the minimal one, so $\mathrm{Aut}(\mathsf{X}_V, \mathbb{Z}^2) = \mathcal{G} \cong \mathbb{Z}^2$, while the extended symmetry group is*

$$\mathrm{Sym}(\mathsf{X}_V, \mathbb{Z}^2) \, = \, \mathrm{Aut}(\mathsf{X}_V, \mathbb{Z}^2) \rtimes \mathrm{Aut}(\mathbb{Z}^2) \, \cong \, \mathbb{Z}^2 \rtimes \mathrm{GL}_2(\mathbb{Z}),$$

hence the maximal extension possible. □

This result shows that positive (topological) entropy is very well compatible with a minimal centralizer, while the factor group $\mathrm{Sym}(\mathsf{X}_V, \mathbb{Z}^2)/\,\mathrm{Aut}(\mathsf{X}_V, \mathbb{Z}^2)$ need neither be a finite nor a periodic group, where the latter statement implies that the factor group contains elements of infinite order. This combination can also occur for subshifts with zero entropy, as can be seen from the subshift that is obtained as the orbit closure of a singleton configuration and contains the shift orbit of this configuration together with the all-0 configuration; compare [6, Ex. 4.3].

6.4. General lattice setting

The statement of Corollary 6.4 is not restricted to $d = 2$. Indeed, one has the following generalisation; see [8, 11, 84] for its first part.

Theorem 6.5 *Let $V = \{(n_1, \ldots, n_d) \in \mathbb{Z}^d : \gcd(n_1, \ldots, n_d) = 1\} = \mathbb{Z}^d \setminus \bigcup_p (p\mathbb{Z}^d)$ be the set of visible points of $\mathbb{Z}^d$, with $d \geq 2$, and consider the topological dynamical system $\bigl(\mathsf{X}_V, \mathbb{Z}^d\bigr)$ with $\mathsf{X}_V = \overline{\mathbb{Z}^d + V}$. Then, X_V has topological entropy $\log(2)/\zeta(d)$ and satisfies $\mathsf{X}_V = \mathbb{A}$,* *where $\mathbb{A}$ consists of all admissible subsets of $\mathbb{Z}^d$, that is, all subsets $U \subset \mathbb{Z}^d$ such that, for every $p \in \mathbb{P}$, U misses at least one coset modulo $p\mathbb{Z}^d$.*

The automorphism group, or topological centralizer, of this system is $\mathrm{Aut}(\mathsf{X}_V, \mathbb{Z}^d) = \mathbb{Z}^d$, *while its extended symmetry group, or topological normalizer, is* $\mathrm{Sym}(\mathsf{X}_V, \mathbb{Z}^d) = \mathbb{Z}^d \rtimes \mathrm{GL}_d(\mathbb{Z})$.

Proof. The statement on the centralizer is a rigidity result that is driven by the identity $\mathsf{X}_V = \mathbb{A}$, which also forces X_V to be hereditary. It follows from a slight modification of the argument put forward in [77], which we repeat here in a form that is tailored to the higher-dimensional lattice systems we consider here and below. It employs a lattice version of the Chinese remainder theorem (CRT) based on the pairwise coprime sublattices of the form $p\mathbb{Z}^d$ (p prime) of the integer lattice. Note that the solutions of a system of congruences appear lattice-periodically, which guarantees some flexibility regarding the actual position of solutions in the square lattice. This argument also works for general lattices.

We start from the identity $\mathsf{X}_V = \mathbb{A}$, which follows from Proposition 6.3 together with its generalisation in Proposition 6.7 and Theorem 6.8 below. First, we show that any automorphism $S \in \mathrm{Aut}(\mathsf{X}_V, \mathbb{Z}^d)$ acts on the singleton set $U_0 = \{0\} \in \mathsf{X}_V$ as a translation, that is, $S(U_0) = U_0 + k$ for some $k \in \mathbb{Z}^d$, where $U_0 \in \mathsf{X}_V$ follows from $\mathsf{X}_V = \mathbb{A}$. Since S is a homeomorphism that commutes with the shift action, it corresponds to a block code ϕ, by the CHL theorem. Here and in what follows, we identify any subset of $\mathbb{Z}^d$ with its characteristic function, and thus with a binary configuration, as explained in Section 6.2. Then, $S(U_0) = U_0 + k$ is equivalent to saying that ϕ takes the value 1 on exactly one block with singleton support. For the latter, note first that ϕ cannot take the value 0 on all blocks with singleton support, as this would imply $S(U_0) = \varnothing$ which is impossible (S is invertible and we already have $S(\varnothing) = \varnothing$).

Assuming the existence of two different blocks with singleton support that are sent to 1 by the code, there is a prime p and an admissible set $U \subset \mathbb{Z}^d$ of cardinality $p^d - 1$ that comprises all cosets modulo $p\mathbb{Z}^d$ except the zero coset, together with the property that $S(U)$ shows *all* possible cosets modulo this very $p\mathbb{Z}^d$ and is thus no longer admissible. To see this, p is chosen such that the difference n of the centers of the two blocks (a non-zero element of $\mathbb{Z}^d$) does *not* belong to $p\mathbb{Z}^d$. In fact, by the CRT, the $p^d - 1$ elements of U can be chosen arbitrarily well separated from one another. Then, the assertion follows because $S(U)$ will contain a translate of $U \cup (n + U)$ and, since $S(U)$ is admissible, a translate of this union is contained in V. Consequently, for some $m \in \mathbb{Z}^d$, both $U + m$ and $(U + n) + m$ consist of $p^d - 1$ elements and are equal modulo $p\mathbb{Z}^d$ (both showing all non-zero cosets modulo $p\mathbb{Z}^d$) — a contradiction to $n \neq 0$ modulo $p\mathbb{Z}^d$ from the construction.

After replacing S by $S' := T_k \circ S$, so that $S'(U_0) = U_0$, and slightly enlarging the size of the block code, one can assume that the only block with singleton support that is sent to 1 is the block that has value 1 only at 0. One is then left to show that $S' = \mathrm{id}$. For convenience, we now rename S' by S, and show that $S = \mathrm{id}$.

This follows from the maximality of V together with the crucial observation that $S(U) \subseteq U$ (equivalently $U \subseteq S^{-1}(U)$) for all $U \in \mathsf{X}_V$, due to the properties of the block code for S just established. So, any (automatically admissible) block of 1_U with value 0 at its central position is sent to 0 by the code. This claim can be shown by an argument similar to the one used above. Assume the existence of an admissible block C with value 0 at its center that is sent to 1 by the code. This block then appears in V at a position s with $s \in p\mathbb{Z}^d$ for a suitable p. Again, one can choose a set U of $p^d - 1$ elements of V that shows all cosets except the zero coset modulo $p\mathbb{Z}^d$. By the CRT, we may assume that these $p^d - 1$ elements are well separated and also well separated from s (together with the whole block $s + C$ of V at s). It is then immediate that $U \cup (s + C)$ is admissible and that $S(U \cup (s + C))$ will contain the set $U \cup \{s\}$ and thus shows all cosets modulo $p\mathbb{Z}^d$, a contradiction.

It remains to determine the normalizer. Since $\mathcal{G} \cong \mathbb{Z}^d$, with $\mathrm{Aut}(\mathbb{Z}^d) = \mathrm{GL}_d(\mathbb{Z})$, there is a group homomorphism

$$\psi \colon \, \mathrm{Sym}(\mathsf{X}_V, \mathbb{Z}^d) \longrightarrow \mathrm{GL}_d(\mathbb{Z})$$

that is induced as follows. If $H \in \mathrm{Sym}(\mathsf{X}_V, \mathbb{Z}^d)$, we have $H\mathcal{G}H^{-1} = \mathcal{G}$, so a set of generators of $\mathcal{G}$ must be mapped to a (possibly different) set of generators under the conjugation action. Starting from our canonical choice, $\mathcal{G} = \langle T_{e_1}, \ldots, T_{e_d} \rangle$, one finds $HT_iH^{-1} = \prod_j T_j^{m_{ji}}$ where the m_{ji} are the matrix elements of $M_H = \psi(H)$. It is routine to verify the homomorphism property. In particular, with $T_n = T_{e_1}^{n_1} \cdots T_{e_d}^{n_d}$, one gets

$$HT_nH^{-1} \, = \, T_{M_H n}. \tag{6.4}$$

For $M \in \mathrm{GL}_d(\mathbb{Z})$, in line with Eq. (6.2), define the mapping H_M on $X = \{0,1\}^{\mathbb{Z}^d}$ by

$$(H_M x)_n \, = \, x_{M^{-1}n},$$

which clearly is a homeomorphism of X. Now, each M maps our set V onto itself, as $\mathrm{GL}_d(\mathbb{Z})$ acts transitively on V. Consequently, also the orbit $\{t + V : t \in \mathbb{Z}^d\}$ is mapped onto itself by M, hence M preserves X_V by continuity. In other words, invoking (6.3), we see that H_M is an element of $\mathrm{Sym}(\mathsf{X}_V, \mathbb{Z}^d)$, and that

$$1 \longrightarrow \mathrm{Aut}(\mathsf{X}_V, \mathbb{Z}^d) \xrightarrow{\mathrm{id}} \mathrm{Sym}(\mathsf{X}_V, \mathbb{Z}^d) \xrightarrow{\psi} \mathrm{GL}_d(\mathbb{Z}) \longrightarrow 1$$

is a short exact sequence. Moreover, the mapping

$$\varphi \colon \, \mathrm{GL}_d(\mathbb{Z}) \longrightarrow \mathrm{Homeo}(\mathsf{X}_V)$$

defined by $\varphi(M) = H_M$ is a group homomorphism as well, with $\psi(H_M) = M$. Consequently, $\mathcal{H} := \varphi(\mathrm{GL}_d(\mathbb{Z}))$ is a subgroup of $\mathrm{Sym}(\mathsf{X}_V, \mathbb{Z}^d)$ that is isomorphic with $\mathrm{GL}_d(\mathbb{Z})$. Since $\varphi \circ \psi$ acts as the identity on $\mathcal{H}$, our claim follows.

Remark With respect to the patch frequency (or Mirsky) measure, the situation is also the same as for $d = 2$, meaning that the dynamical spectrum of $(\mathsf{X}_V, \mathbb{Z}^d, \nu_{\mathrm{M}})$ is pure point, with trivial topological point spectrum. Nevertheless, the measure-theoretic eigenfunctions are continuous on a subset of X_V of full measure; see the discussion in [9].

In fact, the above multi-dimensional setting allows for a further generalisation.

Definition 6.6 *Let $\mathcal{B} = \{b_i \mid i \in \mathbb{N}\}$ be an infinite set of positive integers that is **primitive** in the sense that $b_i \mid b_j$ implies $i = j$. Consider the point set $V_{\mathcal{B}} = \mathbb{Z}^d \setminus \bigcup_{i \in \mathbb{N}} b_i \mathbb{Z}^d$ in $\mathbb{R}^d$, and define $\mathsf{X}_{\mathcal{B}} = \overline{\mathbb{Z}^d + V_{\mathcal{B}}}$, which is compact. Then, the dynamical system $(\mathsf{X}_{\mathcal{B}}, \mathbb{Z}^d)$ is called a **$\mathcal{B}$-free lattice system**. It is called **Erdős** when the b_i are pairwise coprime and satisfy*

$$\sum_{i=0}^{\infty} \frac{1}{b_i^d} \, < \, \infty,$$

which is an additional condition only for $d = 1$.

Note that $d = 1$ is the case of $\mathcal{B}$-free systems in $\mathbb{Z}$, which is extensively studied in [36, 56] and references therein. The primitivity condition really is some irreducibility notion, as any multiple of some b_i could simply be removed from the set $\mathcal{B}$ without any effect on $V_{\mathcal{B}}$. It is obvious that $\mathbb{Z}^d$ in Definition 6.6 can be replaced by any lattice $\Gamma \subset \mathbb{R}^d$. However, since this does not change the arithmetic situation at hand, we restrict our attention to $\mathbb{Z}^d$ for now.

A set $U \subset \mathbb{Z}^d$ is called **admissible** for $\mathcal{B}$ if, for every $b \in \mathcal{B}$, U meets at most $b^d - 1$ cosets of the sublattice $b\mathbb{Z}^d$. Equivalently, U is admissible if it misses at least one coset of $b\mathbb{Z}^d$ for each $b \in \mathcal{B}$. The set of all admissible subsets of $\mathbb{Z}^d$ is again denoted by $\mathbb{A}$, and constitutes a subshift. By definition, $V_{\mathcal{B}} \in \mathbb{A}$, and we thus have $\mathsf{X}_{\mathcal{B}} \subseteq \mathbb{A}$. If P and Q are disjoint finite subsets of $\mathbb{Z}^d$, we define the **locator set**

$$L(P, Q) := \{t \in \mathbb{Z}^d : t + P \subset V_{\mathcal{B}} \text{ and } t + Q \subset \mathbb{Z}^d \setminus V_{\mathcal{B}}\}$$

in analogy to the treatment in [84]. One has the following connection, which is a generalisation of both [37, Prop. 2.5] and [84, Thm. 2].

Proposition 6.7 *Assume that* $\big(\mathsf{X}_{\mathcal{B}}, \mathbb{Z}^d\big)$ *is Erdős, and let* P *and* Q *be disjoint finite subsets of* $\mathbb{Z}^d$*. Then, the following properties are equivalent.*

1. $L(P, Q)$ *has positive natural density.*
2. $L(P, Q) \neq \varnothing$.
3. P *is admissible for* $\mathcal{B}$.

Proof. The implication (1) ⇒ (2) is clear. If $L(P, Q) \neq \varnothing$, one has $t + P \subset V_{\mathcal{B}}$ for some $t \in \mathbb{Z}^d$, so $t + P \in \mathbb{A}$ and hence $P \in \mathbb{A}$, which shows (2) ⇒ (3).

It remains to prove (3) ⇒ (1). To this end, let $m = |P|$ and set

$$S_1 := \{b \in \mathcal{B} : \text{mín}\big(|P \text{ mód } b|, b^d - 1\big) < m\},$$

which is a finite subset of $\mathcal{B}$. Further, for the elements $q \in Q$, select distinct elements b_q from $\mathcal{B} \setminus S_1$, and set $S_2 = \{b_q : q \in Q\}$. Without loss of generality, we may choose each b_q large enough so that $p \equiv q \text{ mód } b_q$ has no solution with $p \in P$, which is to say that q is a representative of a coset modulo b_q that is missed by P. Since $|S_2| = |Q|$, $S := S_1 \cup S_2$ is still a finite subset of $\mathcal{B}$, with $S = S_1$ for $Q = \varnothing$.

Since P is admissible for $\mathcal{B}$, we know that, for each $b \in \mathcal{B}$, at least one coset of $b\mathbb{Z}^d$ is missed by P. Let p_b be a representative of this coset, where we may choose $p_b = q$ for all $b = b_q \in S_2$ due to our choice of S_2. As our system is Erdős, we can invoke the lattice version of the CRT to see that there is an element $t_0 \in \mathbb{Z}^d$ such that

$$t_0 \equiv -p_b \text{ mód } b, \quad \text{for all } b \in S.$$

Note that, with the choice of the p_b for $b \in S_2$ just made, this comprises the congruences $t_0 \equiv -q \text{ mód } b_q$ for all $q \in Q$. In fact, due to the pairwise coprimality, we know that the set of *all* solutions is given by the lattice coset $t_0 + c\mathbb{Z}^d$ with $c = \prod_{b \in S} b$. For any t from this coset and then every $b \in S$, we thus have $t + p \not\equiv 0 \text{ mód } b$, which is to say that $t + P$ avoids the zero coset for all $b \in S$, while $t + q \equiv 0 \text{ mód } b_q$, so no element of $t + Q$ can lie in $V_{\mathcal{B}}$.

Now, let $R_n := \{b \in \mathcal{B} \setminus S : b \leq n\}$, which is finite, where we assume the integer n to be large enough so that $R_n \neq \varnothing$. Now, consider

$$\Theta_n := \bigl(t_0 + c\mathbb{Z}^d\bigr) \cap \{t \in \mathbb{Z}^d : t \not\equiv -p \text{ mód } b \text{ for all } b \in R_n \text{ and all } p \in P\}.$$

The second set is a finite union of translates of the lattice $\gamma_n \mathbb{Z}^d$ with $\gamma_n = \prod_{b \in R_n} b$. Invoking Fact 6.1, it is clear that Θ_n consists of finitely many cosets of the intersection lattice, which is $c\gamma_n \mathbb{Z}^d$, and thus has a well-defined natural density. Consequently, Θ_n has density

$$\operatorname{dens}(\Theta_n) \;=\; c^{-d} \prod_{b \in R_n} \Bigl(1 - \frac{|P|}{b^d}\Bigr)$$

because, modulo b for any $b \in R_n$, no two points of P can be equal by our choice of S_1.

Each term in the product is a positive number, again due to our choice of $S_1 \subseteq S$, so the Erdős condition guarantees that the infinite product satisfies

$$\prod_{b \in \mathcal{B} \setminus S} \Bigl(1 - \frac{|P|}{b^d}\Bigr) \;=\; D \;>\; 0,$$

which is to say that it converges to a positive number. Since $\Theta_{n+1} \subseteq \Theta_n$ for all large enough n, say $n \geq n_0$, we can take the limit $n \to \infty$ and conclude that $\Theta_\infty := \bigcap_{n \geq n_0} \Theta_n$ is a set of solutions of our congruence conditions, for all $b \in \mathcal{B}$, with positive natural density. So, for any $t \in \Theta_\infty$, we have $t + P \subset V_{\mathcal{B}}$ together with $t + Q \subset \mathbb{Z}^d \setminus V_{\mathcal{B}}$ as claimed.

Theorem 6.8 *Let* $\bigl(\mathsf{X}_{\mathcal{B}}, \mathbb{Z}^d\bigr)$ *be a* $\mathcal{B}$*-free lattice system, with* $\operatorname{Aut}(\mathsf{X}_{\mathcal{B}}, \mathbb{Z}^d)$ *its automorphism group. Then, the group of extended symmetries is given by* $\operatorname{Sym}(\mathsf{X}_{\mathcal{B}}, \mathbb{Z}^d) = \operatorname{Aut}(\mathsf{X}_{\mathcal{B}}, \mathbb{Z}^d) \rtimes \mathrm{GL}_d(\mathbb{Z})$*, which is to say that the extension is always the maximally possible one.*

If $\bigl(\mathsf{X}_{\mathcal{B}}, \mathbb{Z}^d\bigr)$ *is Erdős, one has* $\mathsf{X}_{\mathcal{B}} = \mathbb{A}$*, the system is hereditary, and it has minimal automorphism group,* $\operatorname{Aut}(\mathsf{X}_{\mathcal{B}}, \mathbb{Z}^d) = \mathcal{G} \cong \mathbb{Z}^d$*, and we thus get* $\operatorname{Sym}(\mathsf{X}_{\mathcal{B}}, \mathbb{Z}^d) = \mathbb{Z}^d \rtimes \mathrm{GL}_d(\mathbb{Z})$*.*

Proof. Due to the assumptions, any $\mathcal{B}$-free lattice system defines a shift, with faithful shift action, wherefore its automorphism group, $\operatorname{Aut}(\mathsf{X}_{\mathcal{B}}, \mathbb{Z}^d)$, contains a normal subgroup that is isomorphic with $\mathbb{Z}^d$, namely the one generated by the shift action itself, $\mathcal{G}$.

Since $\operatorname{Aut}(\mathbb{Z}^d) = \mathrm{GL}_d(\mathbb{Z})$, any $M \in \mathrm{GL}_d(\mathbb{Z})$ maps $\mathbb{Z}^d$ onto itself, hence one also has $M(b\mathbb{Z}^d) = bM(\mathbb{Z}^d) = b\mathbb{Z}^d$ for any $b \in \mathcal{B}$. This implies $M(V_{\mathcal{B}}) = V_{\mathcal{B}}$. We thus see that $\mathcal{H} := \varphi(\mathrm{GL}_d(\mathbb{Z}))$ is a subgroup of $\operatorname{Sym}(\mathsf{X}_{\mathcal{B}}, \mathbb{Z}^d)$ that is isomorphic with $\mathrm{GL}_d(\mathbb{Z})$. Since we have $\psi(\mathcal{H}) = \psi(\operatorname{Sym}(\mathsf{X}_{\mathcal{B}}, \mathbb{Z}^d)) = \mathrm{GL}_d(\mathbb{Z})$, where ψ is the group homomorphism from above, $\operatorname{Aut}(\mathsf{X}_{\mathcal{B}}, \mathbb{Z}^d)$ is the kernel of the group endomorphism $\varphi \circ \psi$. By construction, $\varphi \circ \psi$ acts as the identity on $\mathcal{H}$, and the claimed semi-direct product structure follows.

Clearly, we have $\mathsf{X}_{\mathcal{B}} \subseteq \mathbb{A}$, as explained earlier. For the converse inclusion, when $\mathsf{X}_{\mathcal{B}}$ is Erdős, consider an arbitrary $S \in \mathbb{A}$ and, for $n \in \mathbb{N}$, set $S_n = S \cap B_n(0)$, which is finite. By Proposition 6.7, for each $n \in \mathbb{N}$, there exists some $t_n \in L\bigl(S_n, (\mathbb{Z}^d \cap B_n(0)) \setminus S_n\bigr) \neq \varnothing$, which means that

$$(V_{\mathcal{B}} - t_n) \cap B_n(0) \;=\; S_n.$$

Consequently, $\text{lím}_{n \to \infty}(V_{\mathcal{B}} - t_n) = S$ in the local topology, and $S \in \mathsf{X}_{\mathcal{B}}$. This shows $\mathbb{A} \subseteq \mathsf{X}_{\mathcal{B}}$ and hence $\mathsf{X}_{\mathcal{B}} = \mathbb{A}$. Clearly, $\mathsf{X}_{\mathcal{B}}$ is then also hereditary. Now, a straight-forward modification of the centralizer argument used in the proof of Theorem 6.5 establishes $\operatorname{Aut}(\mathsf{X}_{\mathcal{B}}, \mathbb{Z}^d) = \mathbb{Z}^d$.

Alternatively, the structure of the last proof can be summarised in stating that

$$1 \longrightarrow \mathrm{Aut}(\mathsf{X}_{\mathcal{B}}, \mathbb{Z}^d) \xrightarrow{\mathrm{id}} \mathrm{Sym}(\mathsf{X}_{\mathcal{B}}, \mathbb{Z}^d) \xrightarrow{\psi} \mathrm{GL}_d(\mathbb{Z}) \longrightarrow 1$$

is a short exact sequence where $\mathcal{H} := \varphi(\mathrm{GL}_d(\mathbb{Z}))$ is a subgroup of $\mathrm{Sym}(\mathsf{X}_{\mathcal{B}}, \mathbb{Z}^d)$ with $\mathcal{H} \cong \mathrm{GL}_d(\mathbb{Z})$ and the property that $\varphi \circ \psi$ acts as the identity on $\mathcal{H}$. Outside the class of Erdős $\mathcal{B}$-free lattice systems, the centralizer can indeed be a finite-index extension of $\mathcal{G}$, as is known from one-dimensional examples of Toeplitz type [58], but we do not consider this case below.

Example Let $k \in \mathbb{N}$ be fixed and consider the lattice $\mathbb{Z}^d$. Then, $\mathcal{B} = \{p^k : p \in \mathbb{P}\}$ leads to the k-free lattice points in d dimensions, which is Erdős for $kd \geq 2$. They have been studied from various angles in [8, 11, 84], and provide a natural extension of our motivating example from Section 6.3.

In particular, one always obtains a measure-theoretic dynamical system $\big(\mathsf{X}_{V_{\mathcal{B}}}, \mathbb{Z}^d, \nu_{\mathrm{M}}\big)$ with pure point diffraction and dynamical spectrum, as in Remark 6.4. The topological entropy is $\log(2)/\zeta(kd)$, while the measure-theoretic entropy with respect to the natural patch frequency (or Mirsky) measure ν_{M} always vanishes [84], as it must in view of the fact that the dynamical spectrum of $\big(\mathsf{X}_{V_{\mathcal{B}}}, \mathbb{Z}^d, \nu_{\mathrm{M}}\big)$ is pure point.

The result of Theorem 6.8 can more generally be looked at as follows. Let $\big(X, \mathbb{Z}^d\big)$ be a faithful shift, with centralizer $\mathrm{Aut}(X, \mathbb{Z}^d)$ and normalizer $\mathrm{Sym}(X, \mathbb{Z}^d)$, and assume that $h_M \in \mathrm{Homeo}(X)$ for some $M \in \mathrm{GL}_d(\mathbb{Z})$, where h_M is the mapping defined in Eq. (6.2). Let T_n with $n \in \mathbb{Z}^d$ denote the shift by n as before, so $\big(T_n x\big)_m = x_{m+n}$, and consider an element $H \in \mathrm{Sym}(X, \mathbb{Z}^d)$ with $M = \psi(H)$. Then, for any $\ell \in \mathbb{Z}^d$, one obtains the commutative diagram

$$\begin{tikzcd}
X \arrow[r, "H"] \arrow[d, "T_\ell"'] & X \arrow[r, "h_{M^{-1}}"] \arrow[d, "T_{M\ell}"'] & X \arrow[d, "T_\ell"] \\
X \arrow[r, "H"] & X \arrow[r, "h_{M^{-1}}"] & X
\end{tikzcd} \qquad (6.5)$$

from Eq. (6.4), where $h_{M^{-1}} \in \mathrm{Homeo}(X)$ by assumption. In particular, $h_{M^{-1}} \circ H \in \mathrm{Homeo}(X)$ commutes with the shift action, hence is a block map by the CHL theorem.

At this point, the structure of the centralizer enters crucially, and one obtains an interesting consequence as follows, where $\psi\colon\ \mathrm{Sym}(X, \mathbb{Z}^d) \longrightarrow \mathrm{Aut}(\mathbb{Z}^d)$ is the homomorphism from above.

Corollary 6.9 *Let $\big(X, \mathbb{Z}^d\big)$ be a faithful subshift with trivial centralizer. Consider an element $H \in \mathrm{Sym}(X, \mathbb{Z}^d)$ with $h_{\psi(H)} \in \mathrm{Homeo}(X)$. Then, H is an affine mapping and $h_{\psi(H)} \in \mathrm{Sym}(X, \mathbb{Z}^d)$.*

Proof. From the diagram (6.5), with $M = \psi(H)$, we know that $h_{M^{-1}} \circ H \in \mathrm{Aut}(X, \mathbb{Z}^d)$, so this mapping equals T_n for some $n \in \mathbb{Z}^d$. This means $H = h_M \circ T_n$, which acts as

$$(Hx)_m = x_{M^{-1}m+n}.$$

The equivalent formulation with sets, due to the relation $h_M \circ T_n = T_{Mn} \circ h_M$, now reads $H(U) = -Mn + M(U)$, which is affine.

Finally, since $H \in \mathrm{Sym}(X, \mathbb{Z}^d)$, one also has $h_M = H \circ T_{-n} \in \mathrm{Sym}(X, \mathbb{Z}^d)$.

The occurrence of affine mappings in the context of $\mathbb{Z}^d$-actions, as a sign of some degree of rigidity, is also known from [63, Thm. 1.1], and will become important later.

6.5. Number theoretic setting

The concept of a $\mathcal{B}$-free lattice system from Definition 6.6 is only one possibility to generalise the one-dimensional notion. For another, combining methods from the theory of aperiodic order [6] with classic results from elementary and algebraic number theory [17, 81], one may start with the treatment of square-free integers in algebraic number fields as in [25], and simplify and generalise it as follows.

Let $\mathbb{K}$ be an algebraic number field of degree d, so $[\mathbb{K} : \mathbb{Q}] = d < \infty$. Let $\mathcal{O}_{\mathbb{K}}$ be the ring of integers in $\mathbb{K}$, which is the unique maximal order in $\mathbb{K}$, such as $\mathbb{Z}$ for $\mathbb{K} = \mathbb{Q}$, $\mathbb{Z}[\mathrm{i}]$ for $\mathbb{K} = \mathbb{Q}(\mathrm{i})$, or $\mathbb{Z}[\sqrt{2}\,]$ for $\mathbb{K} = \mathbb{Q}(\sqrt{2}\,)$. Let $\iota \colon \mathcal{O}_{\mathbb{K}} \longrightarrow \mathbb{R}^r \times \mathbb{C}^s$ be the mapping defined by

$$z \mapsto \big(\rho_1(z), \ldots, \rho_r(z), \sigma_1(z), \ldots, \sigma_s(z)\big),$$

where $\rho_1, \ldots, \rho_r$ are the real embeddings of $\mathbb{K}$ into $\mathbb{C}$, while $\sigma_1, \ldots, \sigma_s$ arise from the complex embeddings of $\mathbb{K}$ into $\mathbb{C}$ by choosing exactly one embedding from each pair of complex conjugate ones (in particular, we have $d = r + 2s$). Clearly, depending on $\mathbb{K}$, one either takes ρ_1 or σ_1 to be the identity.

Now, if $\mathfrak{b}$ is a non-zero ideal of $\mathcal{O}_{\mathbb{K}}$, its absolute **norm** is defined by $N_{\mathbb{K}|\mathbb{Q}}(\mathfrak{b}) := [\mathcal{O}_{\mathbb{K}} : \mathfrak{b}]$; see Chapter 2. In fact, for any of the above choices, the image $\iota(\mathfrak{b})$ is a lattice in $\mathbb{R}^r \times \mathbb{C}^s \cong \mathbb{R}^d$, and the absolute norm of $\mathfrak{b}$ is precisely the index of the sublattice $\iota(\mathfrak{b})$ in the lattice $\iota(\mathcal{O}_{\mathbb{K}})$, and thus a finite number. The map ι is usually called the **Minkowski embedding** of $\mathcal{O}_{\mathbb{K}}$; see [6, 17, 81] for details.

To continue, let $\mathbb{K}$ be an algebraic number field of degree d, and $\mathcal{O}_{\mathbb{K}}$ its ring of integers, with Minkowski embedding $\Gamma = \iota(\mathcal{O}_{\mathbb{K}}) \subset \mathbb{R}^d$. Let $\mathcal{B} = \{\mathfrak{b}_i \mid i \in \mathbb{N}\}$ be an infinite set of non-trivial ideals of $\mathcal{O}_{\mathbb{K}}$, where $\mathcal{B}$ is assumed to be **primitive** in the sense that $\mathfrak{b}_i \supseteq \mathfrak{b}_j$ implies $i = j$. Let $\Gamma_i = \iota(\mathfrak{b}_i)$ and consider $V_{\mathcal{B}} := \Gamma \setminus \bigcup_{i\in\mathbb{N}} \Gamma_i \subset \mathbb{R}^d$, which thus is the Minkowski embedding of $\mathcal{O}_{\mathbb{K}} \setminus \bigcup_{i\in\mathbb{N}} \mathfrak{b}_i$, and define its hull as the orbit closure

$$\mathsf{X}_{\mathcal{B}} \, = \, \overline{\Gamma + V_{\mathcal{B}}}$$

in the local topology, so $\mathsf{X}_{\mathcal{B}}$ is compact as in our previous examples.

Definition 6.10 *In the setting just explained, the topological dynamical system* $\big(\mathsf{X}_{\mathcal{B}}, \Gamma\big)$ *is called an* ***algebraic*** $\boldsymbol{\mathcal{B}}$***-free lattice system****, or simply an* ***algebraic*** $\boldsymbol{\mathcal{B}}$***-free system****.*

Such a system is called ***Erdős*** *when the* $\mathfrak{b}_i$ *are pairwise coprime (meaning* $\mathfrak{b}_i + \mathfrak{b}_j = \mathcal{O}_{\mathbb{K}}$ *for all* $i \neq j$*) and satisfy*

$$\sum_{i=1}^{\infty} \frac{1}{N_{\mathbb{K}|\mathbb{Q}}(\mathfrak{b}_i)} \, < \, \infty .$$

As before, we call a set $U \subset \Gamma$ **admissible** for $\mathcal{B}$ when, for every $\mathfrak{b} \in \mathcal{B}$, the set U meets at most $N_{\mathbb{K}|\mathbb{Q}}(\mathfrak{b}) - 1$ cosets of $\Gamma_{\mathfrak{b}} := \iota(\mathfrak{b})$ in Γ, that is, misses at least one. All admissible subsets of Γ once again constitute a subshift, denoted by $\mathbb{A}$, which contains $\mathsf{X}_{\mathcal{B}}$ by construction.

Proposition 6.11 *Assume that* $\big(\mathsf{X}_{\mathcal{B}}, \Gamma\big)$ *is Erdős, and let* P *and* Q *be disjoint finite subsets of* Γ*. Then, the following properties are equivalent.*

1. *The locator set* $L(P,Q) := \{t \in \Gamma : t + P \subset V_{\mathcal{B}}$ *and* $t + Q \subset \Gamma \setminus V_{\mathcal{B}}\}$ *has positive natural density.*
2. $L(P,Q) \neq \varnothing$.
3. P *is admissible for* $\mathcal{B}$.

Proof. This is a variant of the proof of Proposition 6.7, where (1) ⇒ (2) ⇒ (3) is again clear. We thus need to establish (3) ⇒ (1).

Let $m = |P|$ and choose S_1 as the set of all ideals $\mathfrak{b} \in \mathcal{B}$ such that $|P \text{ mód } \Gamma_{\mathfrak{b}}| < m$ or $N_{\mathbb{K}|\mathbb{Q}}(\mathfrak{b}) \leq m$. As before, S_1 is finite. Then, for S_2, select distinct ideals from $\mathcal{B} \setminus S_1$, denoted by $S_2 = \{\mathfrak{b}_q : q \in Q\}$, where, without loss of generality, we may select ideals $\mathfrak{b}_q$ of sufficiently large absolute norm such that P does not meet the coset modulo $\mathfrak{b}_q$ represented by q. Then, consider $S = S_1 \cup S_2$, which is still finite. As all ideals $\mathfrak{b} \in \mathcal{B}$ can be viewed as lattices $\Gamma_{\mathfrak{b}}$ via the Minkowski embedding, we can again invoke the CRT to find an element $t_0 \in \Gamma$ so that

$$t_0 \equiv -p_{\mathfrak{b}} \text{ mód } \Gamma_{\mathfrak{b}}, \quad \text{for all } \mathfrak{b} \in S,$$

where $p_{\mathfrak{b}}$ is a representative of a coset modulo $\Gamma_{\mathfrak{b}}$ that is missing in P, which we know to exist. Due to our construction of S_2, we may choose $p_{\mathfrak{b}_q} = q$ for all $q \in Q$, wherefore the above congruences actually comprise $t_0 \equiv -q \text{ mód } \Gamma_{\mathfrak{b}_q}$ for all $q \in Q$. By pairwise coprimality of the $\mathfrak{b} \in \mathcal{B}$, we see that the set of all solutions is the coset $t_0 + G$, where G is the Minkowski embedding of the ideal $\prod_{\mathfrak{b} \in S} \mathfrak{b}$. For any t from this coset, $t + P$ avoids the zero coset of $\Gamma_{\mathfrak{b}}$ for all $\mathfrak{b} \in S$, while no element of $t + Q$ is in $V_{\mathcal{B}}$, so $t + Q \subset \Gamma \setminus V_{\mathcal{B}}$.

Now, for $n \in \mathbb{N}$, consider the set $R_n := \{\mathfrak{b} \in \mathcal{B} \setminus S : N_{\mathbb{K}|\mathbb{Q}}(\mathfrak{b}) \leq n\}$. For a suitable n_0 and then all $n \geq n_0$, the set R_n is non-empty and finite. Next, define

$$\Theta_n := (t_0 + G) \cap \{t \not\equiv -p \text{ mód } \Gamma_{\mathfrak{b}} \text{ for all } \mathfrak{b} \in R_n \text{ and all } p \in P\}.$$

Then, Θ_n is once again a finite union of translates of a non-trivial intersection lattice and thus a set of positive natural density, the latter being given by

$$\text{dens}(G) \prod_{\mathfrak{b} \in R_n} \Bigl(1 - \frac{|P|}{N_{\mathbb{K}|\mathbb{Q}}(\mathfrak{b})}\Bigr).$$

As in the previous case, the product is convergent as $n \to \infty$ by the Erdős condition, so $\Theta_\infty := \bigcap_{n \geq n_0} \Theta_n$ is a subset of Γ of positive density such that, for any $t \in \Theta_\infty$, we have $t + P \subset V_{\mathcal{B}}$ and $t + Q \subset \Gamma \setminus V_{\mathcal{B}}$.

Theorem 6.12 *An Erdős algebraic* $\mathcal{B}$*-free system* $\bigl(\mathsf{X}_{\mathcal{B}}, \Gamma\bigr)$ *satisfies* $\mathsf{X}_{\mathcal{B}} = \mathbb{A}$ *and is hereditary. Moreover, it has minimal automorphism group, which means* $\text{Aut}(\mathsf{X}_{\mathcal{B}}, \Gamma) = \Gamma \cong \mathbb{Z}^d$*. Moreover, its extended symmetry group is of the form* $\text{Sym}(\mathsf{X}_{\mathcal{B}}, \Gamma) = \text{Aut}(\mathsf{X}_{\mathcal{B}}, \Gamma) \rtimes \mathcal{H}$*, where* $\mathcal{H}$ *is isomorphic to a non-trivial subgroup of* $\text{GL}_d(\mathbb{Z})$*.*

Proof. While $\mathsf{X}_{\mathcal{B}} \subseteq \mathbb{A}$ is clear, $\mathbb{A} \subseteq \mathsf{X}_{\mathcal{B}}$ is shown exactly as in Theorem 6.8, this time on the basis of Proposition 6.11, so $\mathsf{X}_{\mathcal{B}} = \mathbb{A}$, and this shift is hereditary. Then, the statement on the centralizer follows, once again, from a straight-forward modification of the argument used in the proof of Theorem 6.5.

Let $X = \mathsf{X}_{\mathcal{B}}$, and consider an arbitrary $H \in \mathrm{Sym}(X, \Gamma)$. Here, we have $M := \psi(H) \in \mathrm{Aut}(\Gamma)$ in analogy to our previous cases, and the diagram (6.5) changes to

$$\begin{tikzcd}
X \arrow[r, "H"] \arrow[d, "T_\ell"'] & X \arrow[r, "h_{M^{-1}}"] \arrow[d, "T_{M\ell}"'] & Y \arrow[d, "T_\ell"] \\
X \arrow[r, "H"] & X \arrow[r, "h_{M^{-1}}"] & Y
\end{tikzcd} \tag{6.6}$$

where $Y := h_{M^{-1}}(X)$, while T_ℓ with $\ell \in \Gamma$ is the shift in this case. Note that both X and Y are subshifts of $\{0,1\}^{\Gamma}$, on which T_n and h_M are still well defined, and it is clear that $\varnothing \in X \cap Y$. This new diagram is again commutative, so $\chi = h_{M^{-1}} \circ H$ intertwines the shift actions on X and Y. Consequently, by the CHL theorem, χ is a block map.

The space Y inherits important properties from X, such as its characterisation through admissibility (now defined via the images of cosets in X under $h_{M^{-1}}$) as well as being hereditary. After minor modifications, the arguments from the proof of Theorem 6.5 now show that χ must be a shift map, hence equal to T_n for some $n \in \Gamma$. But $T_n X = X$, whence $H \in \mathrm{Homeo}(X)$ now implies

$$Y \, = \, h_{M^{-1}}(X) \, = \, h_{M^{-1}}(HX) \, = \, T_n X \, = \, X,$$

and we are back to the situation of Corollary 6.9. Consequently, H is an affine mapping, with $H = h_M \circ T_n$, and $h_M \in \mathrm{Sym}(X, \Gamma)$. We thus have a short exact sequence

$$1 \longrightarrow \mathrm{Aut}(X, \Gamma) \xrightarrow{\;\mathrm{id}\;} \mathrm{Sym}(X, \Gamma) \xrightarrow{\;\psi\;} \mathcal{H} := \psi\big(\mathrm{Sym}(X, \Gamma)\big) \longrightarrow 1$$

with $\mathcal{H}$ a subgroup of $\mathrm{Aut}(\Gamma)$. In particular, we get $\mathrm{Sym}(X, \Gamma) = \mathrm{Aut}(X, \Gamma) \rtimes \mathcal{H}$ as claimed.

To see that $\mathcal{H}$ is non-trivial, we observe that the unit group $\mathcal{O}_{\mathbb{K}}^{\times}$ is non-trivial (it contains at least the elements ± 1) and, via the Minkowski embedding, isomorphic to a subgroup of $\mathrm{Aut}(\Gamma) \cong \mathrm{GL}_d(\mathbb{Z})$. Each element of $\mathcal{O}_{\mathbb{K}}^{\times}$ maps any ideal $\mathfrak{b}$ onto itself, so the corresponding mapping induced by the Minkowski embedding is a bijection of $V_{\mathcal{B}}$, and thus gives rise to an extended symmetry. Further elements emerge from non-trivial Galois automorphisms of $\mathbb{K}$, such as complex conjugation when $\mathbb{K}$ is a totally complex extension of $\mathbb{Q}$. Consequently, the claim on the nature of $\mathcal{H}$ is clear.

Remark The systems covered by Theorem 6.12 show many similarities with the k-free lattice points discussed earlier. In particular, they have positive topological entropy, which can in principle be determined from their description as weak model sets of maximal density in the sense of [9]. The spectral properties will reflect the comments made in Remark 6.4. We leave details to the interested reader.

Unlike the situation in Theorem 6.8, the group $\mathcal{H}$ will generally *not* be $\mathrm{Aut}(\Gamma) \cong \mathrm{GL}_d(\mathbb{Z})$, as we shall see in Section 6.6 below. In particular, for $M \in \mathrm{Aut}(\Gamma)$ and $\mathfrak{b} \in \mathcal{B}$, it need not be true that $M(\mathfrak{b}) = \mathfrak{b}$ or $M(V_{\mathcal{B}}) = V_{\mathcal{B}}$. The following negative result, obtained via methods from analytic number theory, was pointed out to us by Valentin Blomer [16].

Fact 6.13 *Let $M \in \mathrm{GL}_2(\mathbb{Z}) \setminus \mathrm{O}(2, \mathbb{Z})$. Then, there exist Gaussian primes $\rho \in \mathbb{Z}[\mathrm{i}] \cong \mathbb{Z}^2$ such that a positive proportion of square-free Gaussian integers $\alpha \in \mathbb{Z}[\mathrm{i}]$ satisfies $\rho^2 \mid M\alpha$, and $M\alpha$ is thus not square-free in $\mathbb{Z}[\mathrm{i}]$.* □

As we shall see in the next section, a simpler statement of purely algebraic nature exists, which suffices for our purposes and permits various generalisations.

6.6. Power-free Gaussian and Eisenstein integers

From now on, we shall need some classic results on quadratic number fields, which can all be drawn from [50, Chs. 14 and 15] or from [96]. To keep things simple, we only consider rings of integers that are Euclidean, so that we can easily work with primes and prime factorisation (up to units) rather than with ideals; see [5] for various generalisations in our context.

As an example of an algebraic $\mathcal{B}$-free system that is Erdős, let us view $\mathbb{Z}^2$ as $\mathbb{Z}[\mathrm{i}]$, the ring of Gaussian integers, and consider, for some fixed $2 \leq k \in \mathbb{N}$, the subset of k-free elements (to be defined below). $\mathbb{Z}[\mathrm{i}]$ is the maximal order in the quadratic field $\mathbb{Q}(\mathrm{i})$, and is Euclidean. The **unit group** of $\mathbb{Z}[\mathrm{i}]$ is

$$\mathbb{Z}[\mathrm{i}]^{\times} \, = \, \{1, \mathrm{i}, -1, -\mathrm{i}\} \, \cong \, C_4 .$$

If $\mathbb{P}$ denotes the set of rational primes as before, the Gaussian primes [50, Thm. 252] can be represented by

$$\mathbb{P}_{\mathrm{G}} \, = \, \{1{+}\mathrm{i}\} \cup \{p \in \mathbb{P} : p \equiv 3 \text{ mód } 4\} \cup \{\pi, \bar{\pi} : \pi\bar{\pi} = p \in \mathbb{P} \text{ with } p \equiv 1 \text{ mód } 4\},$$

where $\bar{\cdot}$ is complex conjugation. The three subsets correspond to the ramified prime, where $(1+\mathrm{i})^2 = 2\mathrm{i}$, the inert primes, and the (complex) splitting primes, respectively. Within the last, by slight abuse of notation, we assume one representing pair for each p to be selected, for instance by demanding π to lie in the positive quadrant. This way, the representation of the primes is unique, and prime factorisation works up to units.

Now, for any integer $k \geq 2$, we can define $V_{\mathrm{G}}^{(k)}$ as the set of Gaussian integers that are not divisible by the kth power of any Gaussian prime. This is the set of k-free Gaussian integers. Figure 6.1 contains an illustration of the set $V_{\mathrm{G}}^{(2)}$, which was also used in [25]. We begin with a geometric symmetry consideration of $V_{\mathrm{G}}^{(k)}$ as follows.

Lemma 6.14 *Let $k \geq 2$ be fixed and let $A \colon \mathbb{Z}[\mathrm{i}] \longrightarrow \mathbb{Z}[\mathrm{i}]$ be a $\mathbb{Z}$-linear bijection that maps $V = V_{\mathrm{G}}^{(k)}$ into itself, $A(V) \subseteq V$. Then, A is a bijection of $U = \mathbb{Z}[\mathrm{i}]^{\times}$, and of V as well. As such, it is of the form $A(x) = \varepsilon\sigma(x)$ with $\varepsilon \in U$ and $\sigma \in \{\mathrm{id}, \bar{\cdot}\}$, that is, ε is a unit and σ a field automorphism of $\mathbb{Q}(\mathrm{i})$.*

Together, these mappings form a group, which is the stabilizer of V in $\mathrm{GL}_2(\mathbb{Z})$, denoted by $\mathrm{Stab}(V)$. The latter, for any $k \geq 2$, is the dihedral group $D_4 \cong C_4 \rtimes C_2$ of order 8, which is the symmetry group of the square and as such a maximal finite subgroup of $\mathrm{GL}_2(\mathbb{Z})$.

Proof. Clearly, any A of the form $A(x) = \varepsilon\sigma(x)$ maps units to units, and V onto itself. Conversely, if A preserves U and $A(1) = \varepsilon$, bijectivity of A implies $A(\mathrm{i}) = \mathrm{i}\varepsilon$ or $A(\mathrm{i}) = -\mathrm{i}\varepsilon$, and $\mathbb{Z}$-linearity of A determines the image of any $x \in \mathbb{Z}[\mathrm{i}]$ from here. In the first case, this gives $A(x) = \varepsilon x$, and $A(x) = \varepsilon\bar{x}$ in the second. It thus remains to show that any $\mathbb{Z}$-linear bijection A of $\mathbb{Z}[\mathrm{i}]$ with $A(V) \subseteq V$ must preserve units.

Let us begin with a simple but powerful observation on the coprimality structure of the k-free Gaussian integers. Consider $x \in V$, with $\gcd_{\mathrm{G}}(x, p) = 1$ for every odd rational prime, where the $\gcd_{\mathrm{G}}$ in $\mathbb{Z}[\mathrm{i}]$ is unique up to units. Then, $p^{\ell} x \in V$ for any $1 \leq \ell < k$, hence also $A(p^{k-1}x) = p^{k-1}A(x) \in V$, which implies $\gcd_{\mathrm{G}}(A(x), p) = 1$. This argument cannot be extended to $p = 2 = -\mathrm{i}(1+\mathrm{i})^2$, which is ramified. Nevertheless, we may conclude that

$$A(U) \subseteq U \cup (1+\mathrm{i})U \cup \cdots \cup (1+\mathrm{i})^{k-1}U,$$

where we now need to exclude all but the first set on the right-hand side.

Observe that, when A is a mapping as specified, then so is the mapping A' defined by $A'(x) = \varepsilon A(x)$, for any $\varepsilon \in U$. We may thus assume $A(1) = (1+\mathrm{i})^m$ for some $0 \leq m \leq k-1$ without loss of generality, matched by $A(\mathrm{i}) = \kappa(1+\mathrm{i})^n$ with $\kappa \in U$ and $0 \leq n \leq k-1$. Now, from $\mathbb{Z}$-linearity in conjunction with bijectivity on $\mathbb{Z}[\mathrm{i}]$, we know that $\det(A) = \pm 1$, where

$$\det(A) \, = \, \Im\big(\overline{A(1)}A(\mathrm{i})\big) \, = \, \Im\big(\kappa(1-\mathrm{i})^m(1+\mathrm{i})^n\big).$$

When $n \geq m$, this gives $\det(A) = 2^m \Im\big(\kappa(1+\mathrm{i})^{n-m}\big)$, which cannot be unimodular unless $m = 0$, so $A(1) = 1$ and $\det(A) = \Im\big(\kappa(1+\mathrm{i})^n\big)$.

Observing $(1+\mathrm{i})^2 = 2\mathrm{i}$, an analogous argument now also excludes $n \geq 2$, so $A(\mathrm{i}) = \kappa$ or $A(\mathrm{i}) = \kappa(1+\mathrm{i})$. In the first case, we get $A(\mathrm{i}) = \mathrm{i}$ or $A(\mathrm{i}) = -\mathrm{i}$ from bijectivity, and A is also a bijection on U. When $A(\mathrm{i}) = \kappa(1+\mathrm{i})$, we get $A(1 \pm \mathrm{i}) = A(1) \pm A(\mathrm{i}) = 1 \pm \kappa(1+\mathrm{i})$. Irrespective of which unit κ is, one of the images is an element of norm 5, where the norm refers to the **field norm**[2] of $x \in \mathbb{Q}(\mathrm{i})$, which is defined by $N(x) = x\bar{x}$ as usual. But such a norm value is impossible by our previous coprimality argument, and thus rules out this case.

When $m > n$, a completely analogous chain of arguments gives $n = 0$ and $m = 1$, which is then once again ruled out by the coprimality result. This leaves us with the mappings that preserve U as claimed.

This result has the following immediate consequence, which can be seen as a simplified (and purely algebraic) case of Fact 6.13.

Corollary 6.15 *Let $k \geq 2$ be a fixed integer and $V = V_{\mathrm{G}}^{(k)}$ the set of k-free Gaussian integers. If $A \in \mathrm{GL}_2(\mathbb{Z}) \setminus \mathrm{Stab}(V)$, there exists a Gaussian prime ρ and an element $w \in V$ such that ρ^k divides $A(w)$.*

No such prime can be inert, and it cannot be ramified when k is even.

Proof. By Lemma 6.14, we know that $A \in \mathrm{GL}_2(\mathbb{Z})$ with $A(V) \subseteq V$ must be an element of $\mathrm{Stab}(V)$, from which the first statement is clear.

The matrix A is unimodular modulo p^k for any rational prime p. As such, it cannot change the number of cosets of $p^k\mathbb{Z}^2$, and maps the zero coset onto itself. This rules out the case that $\rho \in \mathbb{P}_{\mathrm{G}}$ is inert.

If $\rho = 1+\mathrm{i}$, we have $N(\rho) = 2$, and the same argument applies to ρ when k is even.

Under the identification of $\mathbb{Z}[\mathrm{i}]$ with $\mathbb{Z}^2$, let us now consider the subshifts

$$\mathsf{X}_{\mathrm{G}}^{(k)} \, := \, \overline{\mathbb{Z}^2 + V_{\mathrm{G}}^{(k)}},$$

[2] Note that the absolute norm of an ideal in $\mathbb{Z}[\mathrm{i}]$, which is always principal, agrees with the field norm of its generating element in this case.

which share many properties with our previous examples. In particular, they once again satisfy $\mathsf{X}_{\mathrm{G}}^{(k)} = \mathbb{A}$, with the appropriate notion for admissibility, and are hereditary. Further, they have pure point spectrum with trivial topological point spectrum, and the sets $V_{\mathrm{G}}^{(k)}$ are generic elements for the corresponding patch frequency (or Mirsky) measure, the latter defined via any averaging sequence of growing balls centered at 0.

Proposition 6.16 *Let $\big(\mathsf{X}_{\mathrm{G}}^{(k)}, \mathbb{Z}^2\big)$ with fixed $k \geq 2$ be the faithful shift generated by the k-free Gaussian integers. Then, its centralizer is trivial, $\mathrm{Aut}(\mathsf{X}_{\mathrm{G}}^{(k)}, \mathbb{Z}^2) = \mathbb{Z}^2$, while the normalizer $\mathrm{Sym}(\mathsf{X}_{\mathrm{G}}^{(k)}, \mathbb{Z}^2)$ consists of affine transformations only. In particular, $\mathrm{Sym}(\mathsf{X}_{\mathrm{G}}^{(k)}, \mathbb{Z}^2)$ contains a subgroup of the form $\mathbb{Z}^2 \rtimes D_4$, where $D_4 = \mathrm{Stab}\big(V_{\mathrm{G}}^{(k)}\big)$ is the group from Lemma* 6.14.

Proof. The claim on the automorphisms is a consequence of our general result in Theorem 6.12, which asserts that the centralizer is trivial, so $\mathrm{Aut}(\mathsf{X}_{\mathrm{G}}^{(k)}, \mathbb{Z}^2) = \mathbb{Z}^2$.

For the extended symmetries, we are once more in the situation of the diagram (6.6) from the proof of Theorem 6.12. Consequently, by Corollary 6.9, each element of the normalizer is an affine mapping, namely an element of the affine lattice group $\mathbb{Z}^2 \rtimes \mathrm{GL}_2(\mathbb{Z})$.

That $\mathbb{Z}^2 \rtimes D_4$ is a subgroup of $\mathrm{Sym}(\mathsf{X}_{\mathrm{G}}^{(k)}, \mathbb{Z}^2)$ follows from Lemma 6.14. Indeed, since the $\mathbb{Z}^2$-orbit of $V_{\mathrm{G}}^{(k)}$ is dense in $\mathsf{X}_{\mathrm{G}}^{(k)}$ by construction and each element of $\mathrm{Sym}(\mathsf{X}_{\mathrm{G}}^{(k)}, \mathbb{Z}^2)$ is continuous, any $M \in D_4$ maps $\mathsf{X}_{\mathrm{G}}^{(k)}$ onto itself, as does any affine mapping (t, M) with $t \in \mathbb{Z}^2$ and $M \in D_4$.

It remains to complete the determination of $\mathrm{Sym}(\mathsf{X}_{\mathrm{G}}^{(k)}, \mathbb{Z}^2)$, which leads to the following result.

Theorem 6.17 *The automorphism group and the extended symmetry group of $\big(\mathsf{X}_{\mathrm{G}}^{(k)}, \mathbb{Z}^2\big)$, with fixed $k \geq 2$, are given by $\mathrm{Aut}(\mathsf{X}_{\mathrm{G}}^{(k)}, \mathbb{Z}^2) = \mathbb{Z}^2$ and $\mathrm{Sym}(\mathsf{X}_{\mathrm{G}}^{(k)}, \mathbb{Z}^2) = \mathrm{Aut}(\mathsf{X}_{\mathrm{G}}^{(k)}, \mathbb{Z}^2) \rtimes D_4$, respectively, where $D_4 = \mathrm{Stab}(V) = C_4 \rtimes C_2$ is the symmetry group of the square, and as such a maximal finite subgroup of $\mathrm{GL}_2(\mathbb{Z})$. In particular, $C_4 \cong \mathbb{Z}[\mathrm{i}]^{\times}$, while C_2 is the group of field automorphisms of $\mathbb{Q}(\mathrm{i})$, generated by complex conjugation.*

Proof. The role of $\mathbb{Z}^2 \rtimes D_4$ is clear from Proposition 6.16. To complete the proof, we need to show that the only $\mathbb{Z}$-linear, bijective mappings of $\mathsf{X}_{\mathrm{G}}^{(k)}$ onto itself are the ones we already know from Lemma 6.14.

As in the case of k-free lattice points, now by Theorem 6.12, we have $\mathsf{X}_{\mathrm{G}}^{(k)} = \mathbb{A}$, where $\mathbb{A}$ is the subshift that consists of all admissible subsets of $V = V_{\mathrm{G}}^{(k)}$. Here, V itself has the property that, for any $\pi \in \mathbb{P}_{\mathrm{G}}$, precisely the zero coset of the principal ideal (π^k) is missing.

To complete the proof, we have to show that no $\mathbb{Z}$-linear bijection of $\mathbb{Z}[\mathrm{i}] \cong \mathbb{Z}^2$ outside of $\mathrm{Stab}(V)$ can map $\mathbb{A}$ into itself. So, let $A \in \mathrm{GL}_2(\mathbb{Z}) \setminus \mathrm{Stab}(V)$. Then, by Corollary 6.15, there is a $\rho \in \mathbb{P}_{\mathrm{G}}$ and an element $w \in V$ such that $\rho^k \mid A(w)$. Set $n = N(\rho)^k$ and $z_1 = w$. We will now choose Gaussian integers $z_2, \ldots, z_n$ such that the set $S = \{z_1, z_2, \ldots, z_n\}$ is admissible while $A(S)$ meets all cosets of the principal ideal (ρ^k) in $\mathbb{Z}[\mathrm{i}]$.

To this end, choose a non-empty, finite set P of Gaussian primes that contains all primes with $N(\pi) < N(\rho)$ but none with $N(\pi) = N(\rho)$. Concretely, when $N(\rho) > 2$, we just take all primes of smaller norm, while we simply choose the inert prime 3 when $\rho = 1 + \mathrm{i}$. In any case, we have $P = \{\pi_1, \ldots, \pi_m\}$ with $m \geq 1$ this way.

Let $\mathcal{L} = (\pi_1^k \cdots \pi_m^k)$, which is a sublattice of $\mathbb{Z}[\mathrm{i}]$ of index $N(\pi_1 \cdots \pi_m)^k$. Since this index is coprime with $n = N(\rho)^k$, we know from Fact 6.1 that $\mathcal{L}$ meets all cosets of $A^{-1}(\rho^k)$, and so does $1 + \mathcal{L}$, as this is just a translate. So, select numbers $z_2, \ldots, z_n \in 1 + \mathcal{L}$ such that $A(z_2), \ldots, A(z_n)$ meet all non-zero cosets of (ρ^k), and set $S := \{z_1, \ldots, z_n\}$, with $z_1 = w$. Clearly, the set $A(S)$ now meets *all* cosets of (ρ^k) and is thus *not* admissible for ρ, so $A(S) \notin \mathbb{A}$. If we can show that S itself is admissible for all Gaussian primes, we are done.

Clearly, S is admissible for all Gaussian primes π with $N(\pi) > N(\rho)$ by cardinality. If S meets all cosets of (ρ^k), each of them must occur precisely once. Then, we modify S via replacing z_2 by $z_2' = z_2 + w$, which reduces the number of cosets in S by one, without reducing the number of cosets in $A(S)$ because w is k-free with $A(w) \equiv 0 \text{ mód } (\rho^k)$.

If ρ is a splitting prime, we also have to check $\bar{\rho}$, which is not an associate but has the same norm. If S meets all cosets of $(\bar{\rho}^k)$, we need to modify one element z_i with $i > 1$ to remove one coset from S. Due to the previous step, we can neither use z_2' nor the other element of S that is now congruent to z_2' modulo (ρ^k). Since $n \geq 4$, there is at least one other element, z_4 say, that can be replaced by $z_4 + w$. The new set S is now admissible for all Gaussian primes of norm at least $N(\rho)$, while $A(S)$ still meets all cosets of (ρ^k) and is thus not in $\mathbb{A}$.

If $\rho \neq 1 + \mathrm{i}$, it remains to see whether S is now also admissible for all π with $N(\pi) < N(\rho)$. By our construction with the lattice $\mathcal{L}$, we know that, modulo (π^k), all z_i are congruent to w, 1 or $1 + w$, so we meet at most 3 cosets. Since $N(\pi)^k \geq 2^k \geq 4$, we are good, and S is admissible for all Gaussian primes, while $A(S)$ is not, and we have the desired contradiction.

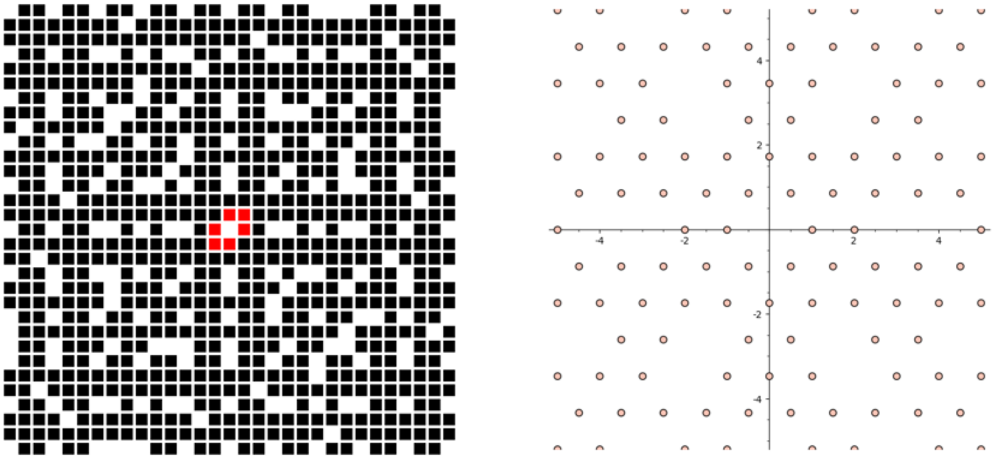


Figure 6.2: The square-free Eisenstein integers, seen both in the $\mathbb{Z}^2$ representation and the corresponding Minkowski embedding. The units are marked in red on the left figure.

A completely analogous chain of arguments works for the ring of Eisenstein integers, $\mathbb{Z}[\rho]$, where $\rho = \mathrm{e}^{2\pi\mathrm{i}/3} = \frac{1}{2}(-1 + \mathrm{i}\sqrt{3}\,)$ is a primitive third root of unity. This is the ring of integers in the imaginary quadratic field $\mathbb{Q}(\rho)$, and is again Euclidean. The unit group is

$$\mathbb{Z}[\rho]^\times \, = \, \{(-\rho)^m : 0 \leq m \leq 5\} \, \cong \, C_6,$$

while the Eisenstein primes [50, Thm. 255], up to units, are represented by

$$\mathbb{P}_{\mathrm{E}} \, = \, \{1 - \rho\} \cup \{p \in \mathbb{P} : p \equiv 2 \text{ mód } 3\} \cup \{\pi, \bar{\pi} : \pi\bar{\pi} = p \in \mathbb{P} \text{ with } p \equiv 1 \text{ mód } 3\},$$

again in the order of the ramified prime, where $(1 - \rho)^2 = -3\rho$, the inert primes, and the complex splitting primes, where one pair $(\pi, \bar{\pi})$ is selected for each p in the last set.

Defining $V_{\mathrm{E}}^{(k)}$ for fixed $k \geq 2$ as the set of k-free Eisenstein integers, which we may either view as a subset of the triangular lattice, which is $\mathbb{Z}[\rho]$, or (equivalently) as one of the square lattice via $\{(m,n) \in \mathbb{Z}^2 : m + n\rho \in V_{\mathrm{E}}^{(k)}\}$, the analogue of Lemma 6.14 now gives mappings of the form $A(x) = \varepsilon\sigma(x)$ with $\varepsilon \in \mathbb{Z}[\rho]^{\times}$ and $\sigma \in \{\mathrm{id}, \bar{\cdot}\}$, hence the group $D_6 \cong C_6 \rtimes C_2$, which is another maximal finite subgroup of $\mathrm{GL}_2(\mathbb{Z})$, this time the one that is the symmetry group of the regular hexagon.

Defining the subshifts

$$\mathsf{X}_{\mathrm{E}}^{(k)} := \overline{\mathbb{Z}[\rho] + V_{\mathrm{E}}^{(k)}},$$

one obtains the following analogue of Theorem 6.17, the proof of which need not be repeated, as the method is the same.

Theorem 6.18 *The automorphism group and the extended symmetry group of* $\bigl(\mathsf{X}_{\mathrm{E}}^{(k)}, \mathbb{Z}[\rho]\bigr)$, *with fixed* $k \geq 2$, *are given by:*

$$\begin{aligned}\mathrm{Aut}(\mathsf{X}_{\mathrm{E}}^{(k)}, \mathbb{Z}[\rho]) &= \mathbb{Z}[\rho] \cong \mathbb{Z}^2,\\ \mathrm{Sym}(\mathsf{X}_{\mathrm{E}}^{(k)}, \mathbb{Z}[\rho]) &= \mathrm{Aut}(\mathsf{X}_{\mathrm{E}}^{(k)}, \mathbb{Z}[\rho]) \rtimes D_6,\end{aligned}$$

respectively, where $D_6 = C_6 \rtimes C_2$ *is the symmetry group of the regular hexagon, and as such isomorphic to a maximal finite subgroup of* $\mathrm{GL}_2(\mathbb{Z})$. *In particular,* $C_6 = \mathbb{Z}[\rho]^{\times}$, *while* C_2 *is the group of field automorphisms of* $\mathbb{Q}(\rho)$, *generated by complex conjugation.* □

So far, we have seen extension groups that are either all of $\mathrm{GL}_2(\mathbb{Z})$ (for the visible lattice points), or finite subgroups thereof (for the k-free Gaussian or Eisenstein integers). In particular, the subshifts defined by the two examples illustrated in Figure 6.1 are clearly distinguished by different extended symmetry groups. At this point, it is a natural question whether also infinite true subgroups of $\mathrm{GL}_2(\mathbb{Z})$ may occur. To this end, we take a look at the corresponding dynamical systems for *real* quadratic fields.

6.7. Power-free integers in real quadratic number fields

Let us first consider subsets of $\mathbb{Z}^2$ constructed by means of k-free integers in $\mathbb{Z}[\sqrt{2}\,]$, namely

$$V_2^{(k)} := \bigl\{(m,n) \in \mathbb{Z}^2 : m + n\sqrt{2} \text{ is } k\text{-free in } \mathbb{Z}[\sqrt{2}\,]\bigr\},$$

where $k \in \mathbb{N}$ with $k \geq 2$ is fixed. This set emerges via the isomorphism between $\mathbb{Z}^2$ and the Minkowski embedding of $\mathbb{Z}[\sqrt{2}\,]$ into $\mathbb{R}^2$; compare [6, Sec. 3.4.1]. Here, with $\lambda := 1 + \sqrt{2}$ denoting the fundamental unit, the unit group is

$$U \,=\, \mathbb{Z}[\sqrt{2}\,]^{\times} \,=\, \{\pm\lambda^n : n \in \mathbb{Z}\} \,\cong\, C_2 \times C_{\infty},$$

where we also note that $\mathbb{Z}[\sqrt{2}\,] = \mathbb{Z}[\lambda]$. This ring is again Euclidean, so we can work with unique prime decomposition up to units.

The primes [50, Thm. 256] can be represented as

$$\mathbb{P}_2 \,=\, \{\sqrt{2}\,\} \cup \{p \in \mathbb{P} : p \equiv \pm 3 \text{ mód } 8\} \cup \{\pi, \pi^{\star} : \pi\pi^{\star} = p \in \mathbb{P} \text{ with } p \equiv \pm 1 \text{ mód } 8\},$$

where $(\cdot)^\star$ denotes the mapping that is the unique extension of $\sqrt{2} \mapsto -\sqrt{2}$ to a field automorphism of the quadratic field $\mathbb{K} = \mathbb{Q}(\sqrt{2}\,)$. The relevant field norm is then given by $N(x) = xx^\star$, which means $N(m+n\sqrt{2}\,) = m^2-2n^2$ or, equivalently, $N(r+s\lambda) = r^2+2rs-s^2$. Once again, to gain a representation modulo units (integers of norm ±1), one pair is selected in the last set for each p. Note that the field norm can be negative here, wherefore the absolute norm of a principal ideal now is the absolute value of the field norm of a generating element.

For some of the calculations below, it is helpful to express λ^n in terms of λ and 1, for arbitrary $n \in \mathbb{Z}$. Defining the bi-infinite sequence $(c_n)_{n\in\mathbb{Z}}$ by the recursion $c_{n+1} = 2c_n + c_{n-1}$ with initial conditions $c_0 = 0$ and $c_1 = 1$, one obtains the analogue of the Fibonacci numbers for the quadratic field $\mathbb{K}$. In particular, they satisfy $c_{-n} = (-1)^{n+1} c_n$ for all $n \in \mathbb{Z}$, and the first few numbers are

$$\ldots, 29, -12, 5, -2, 1, 0, 1, 2, 5, 12, 29, \ldots$$

The required formula for the units now reads

$$\lambda^n \;=\; c_n \lambda + c_{n-1} \;=\; c_n \sqrt{2} + (c_n + c_{n-1}), \tag{6.7}$$

which holds for all $n \in \mathbb{Z}$, as can easily be checked by induction.

Lemma 6.19 *Let $A\colon \mathbb{Z}[\sqrt{2}\,] \longrightarrow \mathbb{Z}[\sqrt{2}\,]$ be a $\mathbb{Z}$-linear bijection that maps $V = V_2^{(k)}$ into itself, for some fixed integer $k \geq 2$. Then, A is of the form $A(x) = \varepsilon\sigma(x)$ with $\varepsilon \in U$ and $\sigma \in \{\mathrm{id}, (\cdot)^\star\}$, so maps $U = \mathbb{Z}[\sqrt{2}\,]^\times$ onto itself. Together, these mappings form the group* $\mathrm{Stab}(V) = U \rtimes C_2 = C_2 \times (C_\infty \rtimes C_2) = C_2 \times D_\infty$ *of infinite order.*

Proof. Any A of the form $A(x) = \varepsilon\sigma(x)$ satisfies $A(V) = V$ and maps U onto itself, while the converse direction will be a consequence of showing that no further $\mathbb{Z}$-linear bijection of $\mathbb{Z}[\sqrt{2}\,]$ exists that maps the set V into itself.

So, let A be a $\mathbb{Z}$-linear bijection of $\mathbb{Z}[\sqrt{2}\,]$ with $A(V) \subseteq V$. As in the proof of Lemma 6.14, we observe that $x \in V$ with $\gcd_{\mathbb{K}}(x, p) = 1$ for any odd $p \in \mathbb{P}$ implies $p^{k-1}x \in V$ and $A(p^{k-1}x) = p^{k-1}A(x) \in V$, hence $\gcd_{\mathbb{K}}(A(x), p) = 1$ as well. Since $2 = (\sqrt{2}\,)^2$, which is the only ramified prime in this case, we see that $A(1)$ and $A(\sqrt{2}\,)$ must be elements of the union

$$U \cup \sqrt{2}U \cup 2U \cup \ldots \cup (\sqrt{2}\,)^{k-1}U,$$

where we may assume that we have, once again without loss of generality, $A(1) = 2^{m/2}$ and $A(\sqrt{2}\,) = \kappa 2^{n/2}$ with $\kappa \in U$ and $0 \leq m, n \leq k-1$. Here, we also know that this must result in a mapping with determinant ±1.

Now, define $W\colon \mathbb{Q}(\sqrt{2}\,) \longrightarrow \mathbb{Q}$ by $W(x) = \bigl(x - x^\star\bigr)/2\sqrt{2}$, and observe that this gives $\det(A) = W\bigl(A(1)^\star A(\sqrt{2}\,)\bigr)$, hence

$$\det(A) \;=\; W\bigl(\kappa(-1)^m(\sqrt{2}\,)^{m+n}\bigr).$$

When $m+n$ is even, so $m+n = 2\ell$, this means $\det(A) = (-1)^m 2^\ell W(\kappa)$, which can only be unimodular if $\ell = 0$ and thus $m = n = 0$. With $\kappa = \pm\lambda^r = \pm\bigl(c_r\sqrt{2} + (c_r + c_{r-1})\bigr)$ from (6.7), we then get $\det(A) = \pm c_r$, which in turn implies $c_r = 1$ and thus $r = \pm1$. So, we have to consider $A(1) = 1$ together with $A(\sqrt{2}\,) = \pm\lambda$. Both choices, however, lead to a contradiction to our coprimality condition by observing that $2 \pm \sqrt{2}$, which has norm 2, is then mapped under A to $3 + \sqrt{2}$, which is a number of norm 7.

Likewise, when $m+n = 2\ell+1$, we have $\det(A) = (-1)^m 2^\ell W(\kappa\sqrt{2}\,)$, which forces $\ell = 0$ and thus either $m = 1$ and $n = 0$ or $m = 0$ and $n = 1$. In both cases, $\kappa = \pm\lambda^r$ can only lead to a unimodular determinant when $c_r + c_{r-1} \in \{\pm 1\}$, which means $r \in \{-1, 0, 1\}$. When $m = 1$ and $n = 0$, we get $A(1) = \sqrt{2}$ together with $A(\sqrt{2}\,) = \kappa \in \{\pm\lambda, \pm 1, \pm\lambda^\star\}$. All six choices lead to contradictions to coprimality with odd primes, by considering images of $1 \pm \sqrt{2}$ or $2 \pm \sqrt{2}$ under A.

It remains to consider $m = 0$ and $n = 1$, so $A(1) = 1$ together with $A(\sqrt{2}\,) = \kappa\sqrt{2}$, with the same options for κ as in the previous case. Once again, $\kappa = \pm\lambda$ and $\kappa = \pm\lambda^\star$ are impossible, as can be seen by considering $A(1 \pm \sqrt{2}\,)$. The choices $\kappa = \pm 1$, however, give the mappings $A(x) = x$ and $A(x) = x^\star$, which map U onto itself, as does any multiplication of such an A with an arbitrary $\varepsilon \in U$.

Let us now consider the subshifts $\mathsf{X}_2^{(k)} := \overline{\mathbb{Z}^2 + V_2^{(k)}}$, in complete analogy to above.

Proposition 6.20 *The automorphism group and the extended symmetry group of* $\bigl(\mathsf{X}_2^{(k)}, \mathbb{Z}^2\bigr)$, *with fixed* $k \geq 2$, *are given by* $\mathrm{Aut}(\mathsf{X}_2^{(k)}, \mathbb{Z}^2) = \mathbb{Z}^2$ *and* $\mathrm{Sym}(\mathsf{X}_2^{(k)}, \mathbb{Z}^2) = \mathrm{Aut}(\mathsf{X}_2^{(k)}, \mathbb{Z}^2) \rtimes \mathcal{H}$, *respectively, where the extension group is* $\mathcal{H} = \mathrm{Stab}\bigl(V_2^{(k)}\bigr) = U \rtimes C_2 \cong C_2 \times D_\infty$, *which is infinite.*

Proof. From Lemma 6.19, we see that $\mathbb{Z}^2 \rtimes \mathcal{H}$ is a subgroup of $\mathrm{Sym}(\mathsf{X}_2^{(k)}, \mathbb{Z}^2)$. The latter is a subgroup of $\mathbb{Z}^2 \rtimes \mathrm{GL}_2(\mathbb{Z})$ by Corollary 6.9. It thus remains to show that $\mathcal{H} = \mathrm{Stab}(V_2^{(k)})$ contains all $\mathrm{GL}_2(\mathbb{Z})$ elements that map $\mathsf{X}_2^{(k)}$ into itself. This last step can be established by the method from the proof of Theorem 6.17, with the field norm replaced by the absolute norm.

There are other real quadratic fields that are Euclidean, such as $\mathbb{Q}(\sqrt{m}\,)$ with $m = 5$ and $m = 3$, which play prominent roles in the theory of aperiodic order, as they are connected with systems with fivefold and twelvefold symmetry, respectively; see [6, Sec. 2.5.1] for background.

For $m = 5$, the ring of integers is $\mathbb{Z}[\tau]$, where $\tau = \frac{1}{2}(1 + \sqrt{5}\,)$ is the golden ratio. Its unit group is $U = \mathbb{Z}[\tau]^\times = \{\pm\tau^n : n \in \mathbb{Z}\}$, and the primes [50, Thm. 257] are represented by

$$\mathbb{P}_5 \;=\; \{\sqrt{5}\,\} \cup \{p \in \mathbb{P} : p \equiv \pm 2 \text{ mód } 5\} \cup \{\pi, \pi^\star : \pi\pi^\star = p \in \mathbb{P} \text{ with } p \equiv \pm 1 \text{ mód } 5\},$$

where $(\cdot)^\star$ is the field automorphism of $\mathbb{Q}(\sqrt{5}\,)$ induced by $\sqrt{5} \mapsto -\sqrt{5}$, with our usual convention for the splitting primes in place. The only ramified prime is 5, while the field norm on $\mathbb{Z}[\tau]$ is $N(m + n\tau) = m^2 + mn - n^2$, which can be negative.

Finally, let us consider the slightly more complicated case $m = 3$, where the ring of integers is $\mathbb{Z}[\sqrt{3}\,]$. Its unit group is given by $\mathbb{Z}[\sqrt{3}\,]^\times = \{\pm\eta^n : n \in \mathbb{Z}\}$, with fundamental unit $\eta = 2 + \sqrt{3}$. Here, in contrast to the two previous cases, all units have norm 1. Employing [96, Thm. 11.1], one sees that the primes up to units can be represented as

$$\begin{aligned}\mathbb{P}_3 \,=\, & \{1 + \sqrt{3}, \sqrt{3}\,\} \cup \{p \in \mathbb{P} : p \equiv \pm 5 \text{ mód } 12\} \\ & \cup \{\pi, \pi^\star : \pi\pi^\star = \pm p \in \mathbb{P} \text{ with } p \equiv \pm 1 \text{ mód } 12\}\end{aligned}$$

with the usual convention for the last set, where $(\cdot)^\star$ is now induced by $\sqrt{3} \mapsto -\sqrt{3}$. Unlike before, since the field discriminant is 12 and thus divisible by 2 and 3, there are *two* ramified primes, where $(1 + \sqrt{3}\,)^2 = 2\eta$ is the additional relation.

This leads to more cases to consider in the determination of $\mathrm{Stab}(V_3^{(k)})$, but the $\mathbb{Z}$-linear bijections of $\mathbb{Z}[\sqrt{3}\,]$ that map $V_3^{(k)}$ into itself, for some fixed $k \geq 2$, are still the expected ones, namely the maps A of the form $A(x) = \varepsilon\sigma(x)$ with $\varepsilon \in U = \mathbb{Z}[\sqrt{3}\,]^{\times}$ and $\sigma \in \{\mathrm{id}, (\cdot)^{\star}\}$; we leave this proof to the interested reader.

In both cases, a proof analogous to the one of Proposition 6.20 gives the following result.

Theorem 6.21 *The automorphism group and the extended symmetry group of* $\bigl(\mathsf{X}_m^{(k)}, \mathbb{Z}^2\bigr)$, *with fixed* $m \in \{2,3,5\}$ *and* $k \geq 2$, *are given by* $\mathrm{Aut}(\mathsf{X}_m^{(k)}, \mathbb{Z}^2) = \mathbb{Z}^2$ *and* $\mathrm{Sym}(\mathsf{X}_m^{(k)}, \mathbb{Z}^2) = \mathrm{Aut}(\mathsf{X}_m^{(k)}, \mathbb{Z}^2) \rtimes \mathcal{H}$, *respectively, where the extension group is* $\mathcal{H} = \mathrm{Stab}\bigl(V_m^{(k)}\bigr) = U \rtimes C_2 \cong C_2 \times D_{\infty}$, *which is an infinite group that does not depend on* k, *where* U *is the unit group as before.* □

The advantage of using the normalizer in addition to the centralizer as a topological invariant becomes obvious in dimensions $d \geq 2$. In [13], this was demonstrated for the chair tiling shift and for Ledrappier's shift. In both cases, $\mathrm{Sym}(X, \mathbb{Z}^2)$ was an extension of $\mathrm{Aut}(X, \mathbb{Z}^2)$ of finite index. As our number-theoretic examples above show, this phenomenon occurs again, but $\mathrm{Sym}(X, \mathbb{Z}^d)$ can also be an *infinite-index* extension of $\mathrm{Aut}(X, \mathbb{Z}^d)$, either for trivial reasons (visible lattice points) or for non-trivial ones (k-free $\mathbb{Z}[\sqrt{2}\,]$-integers). At present, we do not know whether such an infinite extension is also possible for minimal, deterministic (zero entropy) subshifts. In any case, these groups allow the distinction of several subshifts (up to topological conjugacy) that have the same centralizer, but different normalizers, such as $(\mathsf{X}_V, \mathbb{Z}^2)$ and $(\mathsf{X}_{\mathrm{G}}^{(2)}, \mathbb{Z}^2)$ from above.

Due to the nature of the associated dynamical system, the structure of the hull X_V (given the property of hereditariness and the natural topology employed) allows for *local symmetries* (that is, transformations that preserve finite local substructures in the set V up to translation) to manifest themselves. To some extent, they may be observed by analysing the extended symmetry group $\mathrm{Sym}(\mathsf{X}_V, \mathbb{Z}^d)$. While symmetries of the set V in the standard sense (global symmetries) are obviously local symmetries in this new sense, the converse is not clear. It is easy to build sets V that have many local symmetries while lacking global symmetries entirely. Thus, it is interesting to note that, in the current context, those two kind of symmetries happen to be the same, bringing up the question on whether this is a natural phenomenon on sets defined in an 'algebraic' form in more general ways.

This setting deserves further attention, in particular in the context of dynamical systems of number-theoretic origin. As this will require a more general approach via ideals, as well as some additional and less elementary results from algebraic and analytic number theory, we defer this to a separate investigation [5].

6.8. Brief commentary on the general quadratic case and further generalizations

In what follows, we briefly summarize how the examples from the previous section could be generalized to other fields; this is a comment on the not yet published work by Baake, Bustos and Nickel [5]. We start by discussing the general structure of the proofs for the above five examples.

Theorem 6.12 above shows, in a very general setting, that the structure of a $\mathcal{B}$-free shift over a general ring of integers $\mathcal{O}_{\mathbb{K}}$ is closely linked to the collection of all admissible sets; this shows that $\mathsf{X}_{\mathcal{B}}$ always has symmetry rigidity, i.e. $\mathrm{Aut}(\mathsf{X}_{\mathcal{B}}, \Gamma) \cong \Gamma$, where $\Gamma = \mathbb{Z}^{[\mathbb{K}:\mathbb{Q}]}$ is the corresponding Minkowski embedding, and that we may study extended symmetries from their effect on admissible sets. The situation for the extended symmetry group, as seen in the examples above described, is closely related to the generating set $V^{(k)}$ and the collection of matrices which preserve it.

In what follows, we shall identify Γ with $\mathbb{Z}^d$, $d = [\mathbb{K} : \mathbb{Q}]$ via some integral basis, so that the matrices that preserve $V^{(k)}$ belong to $\mathrm{GL}_d(\mathbb{Z})$. In this regard, it is evident that if $A \in \mathrm{GL}_d(\mathbb{Z})$ corresponds to multiplication by a unit, then $A(V^{(k)}) = V^{(k)}$, since multiplication by units preserves (ideal) factorizations. Less evident, but still easily verifiable, is the fact that matrices corresponding to Galois automorphisms also preserve $V^{(k)}$, as they map the lattice corresponding to the ideal $\mathfrak{p}^r$ to itself for any prime ideal $\mathfrak{p}$ which is a factor of an inert or ramified prime, and at most swaps $\mathfrak{p}^r$ with $\bar{\mathfrak{p}}^r$ in the case of a ramified prime, as Galois automorphisms preserve ideal norms; since $V^{(k)}$ is defined by exclusion of such ideals depending only on the exponent, if some element of $\mathbb{Z}^d$ is removed in the definition of this set, its image under A will be removed too. This shows that the quotient group $\mathrm{Sym}(\mathsf{X}_{V^{(k)}}, \Gamma)/\mathrm{Aut}(\mathsf{X}_{V^{(k)}}, \Gamma)$ always contains a copy of $\mathcal{O}_{\mathbb{K}}^{\times} \rtimes \mathrm{Aut}(\mathbb{K}|\mathbb{Q})$, where the latter is the set of all Galois automorphisms of $\mathbb{K}$ over $\mathbb{Q}$.

As seen above in the Gaussian example, the crux of the proof of this inclusion being an equality lies in proving that a matrix A satisfying the condition $A(V^{(k)}) \subseteq V^{(k)}$ necessarily belongs to this copy of $\mathcal{O}_{\mathbb{K}}^{\times} \rtimes \mathrm{Aut}(\mathbb{K}|\mathbb{Q})$. The necessity of proving this property with the reduced hypothesis is that, as a consequence, we may find "bad elements" of $V^{(k)}$ whenever A is a matrix not corresponding to an element of $\mathcal{O}_{\mathbb{K}}^{\times} \rtimes \mathrm{Aut}(\mathbb{K}|\mathbb{Q})$, that is, elements $x \in V^{(k)}$ with $A(x) \notin V^{(k)}$; with analytic number theory arguments, we could argue that these bad elements are actually common (as in, they form a subset of $V^{(k)}$ of positive density), at least in our example cases, but we only need to find one.

The existence of these bad elements, combined with hereditariness, allows us to show that any potential extended symmetry f with associated matrix A can be used to construct a non-admissible set from any admissible finite set $x \in \mathbb{A} = \mathsf{X}_{V^{(k)}}$ using f. Of course, if A represents an element of $\mathcal{O}_{\mathbb{K}}^{\times} \rtimes \mathrm{Aut}(\mathbb{K}|\mathbb{Q})$, such bad elements do not exist, and thus we obtain an isomorphism $\mathrm{Sym}(\mathsf{X}_{V^{(k)}}, \Gamma)/\mathrm{Aut}(\mathsf{X}_{V^{(k)}}, \Gamma) \cong \mathcal{O}_{\mathbb{K}}^{\times} \rtimes \mathrm{Aut}(\mathbb{K}|\mathbb{Q})$, as we wanted to.

Thus, to generalize the above argument to other rings of integers, we need to study the matrices A that send $V^{(k)}$ to some subset of $V^{(k)}$. If we look at the Gaussian example, we see that, for both split and inert primes p, if some $x \in V^{(k)}$ is coprime with p, then $px, p^2x, \dots, p^{k-1}x$ are k-free. This "gcd trick" is part of the analysis we perform on the elements of $V^{(k)}$, to determine if they are "bad" for the matrix A, and this technique allows us to ignore split and inert primes in this situation. To be precise, if we define $W^{(k)}$ as the subset of $V^{(k)}$ comprised only of those elements whose ideal prime factors are ramified, that is:

$$W^{(k)} := \{x \in V^{(k)} \ : \ v_{\mathfrak{p}}(x) = 0 \text{ if } \mathfrak{p} \nmid (d_{\mathbb{K}})\},$$

where $d_{\mathbb{K}}$ is the discriminant of $\mathbb{K}$, we may prove that $A(V^{(k)}) \subseteq V^{(k)}$ implies that $A(W^{(k)}) = W^{(k)}$, and analysis of the latter set is often easier in our situation.

If we restrict ourselves to the case of quadratic fields, the proof of the latter equality differs depending on whether the field is real (i.e. contained in $\mathbb{R}$; see Chapter 2) or imaginary, the

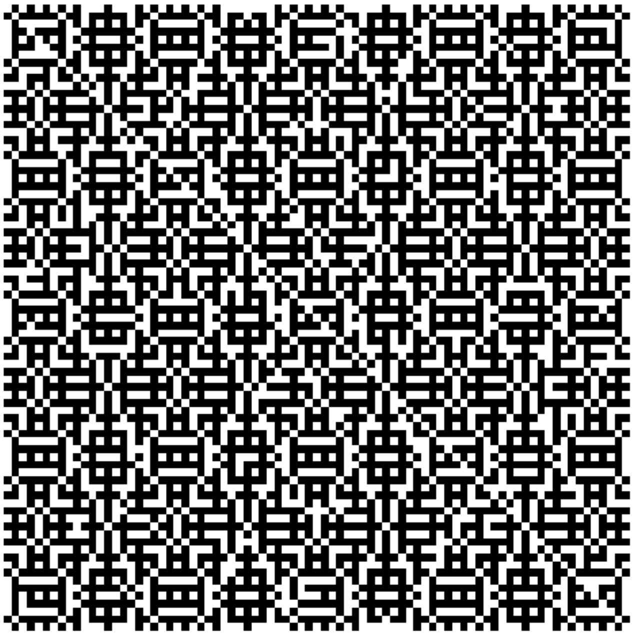


Figure 6.3: The square-free integers in the imaginary quadratic field $\mathbb{Z}[\sqrt{-2}]$.

latter case being quite simple as the set $W^{(k)}$ is then finite (as a consequence of the Dirichlet unit theorem) and thus the obvious inclusion already implies equality. In the real case, $W^{(k)}$ is infinite, but, since the field norms of its element must be divisors of the discriminant $d_{\mathbb{K}}$, it may be seen as the set of points with integer coordinates from a finite union of hyperbolas in the plane $\mathbb{R}^2$, as the field norm of some element $(x, y) \in \mathbb{Z}^2 \cong \mathcal{O}_{\mathbb{K}}$ is a quadratic form on x and y. Thus, we may use the fact that five points determine a conic section (which is a consequence of Bézout's theorem from algebraic geometry; see Chapter 2) to prove that the hyperbolas from this collection are mapped to themselves (up to a permutation), implying then the desired equality.

Staying in the realm of quadratic fields some more, we can prove that a matrix A that preserves $W^{(k)}$ must also preserve the set of units of the integer ring $\mathcal{O}_{\mathbb{K}}$. In the above examples, we use this fact to prove then that A must be a matrix of the desired type as a consequence. For instance, for $\mathbb{Z}[\mathrm{i}]$, we have that $\{1, \mathrm{i}\}$ is an integral basis for the ring and both $A(1)$ and $A(\mathrm{i})$ should be either ± 1 or $\pm\mathrm{i}$; this shows that, since A has determinant ± 1, it must then be one of the eight matrices from the usual 2-dimensional representation of D_4 (see Chapter 1). In the other examples, a similar argument is used, including the usage of determinants in the analysis. For general quadratic fields, this relationship between A and the sets $W^{(k)}$ and $\mathcal{O}_{\mathbb{K}}^{\times}$ is less straightforward to prove, but it can be done; once again the characterization of $\mathcal{O}_{\mathbb{K}}^{\times}$ and $W^{(k)}$ as points with integer coordinates in a collection of hyperbolas proves useful in the real case.

Nonetheless, in both the real and imaginary cases, with some work, we can prove that the equality $A(\mathcal{O}_{\mathbb{K}}^{\times}) = \mathcal{O}_{\mathbb{K}}^{\times}$ is enough to conclude that A must be a matrix of the desired type. The result we obtain with this procedure is as follows:

Theorem 6.22 *Let $k \geq 2$ and $\mathbb{K} = \mathbb{Q}(\sqrt{d})$ be any quadratic field, with d square-free. No* *matter whether $\mathbb{K}$ is real or imaginary, the k-free shift $V^{(k)}$ has extended symmetry group given by*[3]*:*

$$\mathrm{Sym}(\mathsf{X}_{V^{(k)}}, \mathbb{Z}^2) \cong \mathbb{Z}^2 \rtimes (\mathcal{O}_{\mathbb{K}}^{\times} \rtimes \mathrm{Gal}(\mathbb{K}|\mathbb{Q})).$$

This also suggests a method to generalize the above results to other fields: the set $W^{(k)}$ and the corresponding set of units may be characterized by the field norm $N_{\mathbb{K}|\mathbb{Q}}$, which, given a fixed integer basis for $\mathbb{K}$, is given by a polynomial on its coefficients. This brings up algebraic geometry to our study: the set $N_{\mathbb{K}|\mathbb{Q}}(x) = c$ is an affine variety in $\mathbb{R}^{[\mathbb{K}:\mathbb{Q}]}$. Thus, a similar argument to the one used for hyperbolas above can be used to show that $A(V^{(k)}) \subseteq V^{(k)}$ implies $A(W^{(k)}) = W^{(k)}$, using tools such as the decomposition of $N_{\mathbb{K}|\mathbb{Q}}(x) = c$ into irreducible varieties and the Ax–Grothendieck theorem (see Theorem 2.84 at the end of Chapter 2) in place of the five-point characterization for conic sections. As an example on how these arguments can be generalized, note that, as $W^{(k)}$ is the set of integer points of a finite union of irreducible varieties, and any linear map is continuous in the Zariski topology, the equality $AW^{(k)} = W^{(k)}$ implies that if V is one of the aforementioned irreducible varieties, $A(V)$ is one of them as well, and thus A at most induces a permutation between these components. In particular, some power of A must map $\mathcal{O}_{\mathbb{K}}^{\times}$ to itself, even if it is no longer obvious that A itself does so.

Thus, the relationship between $W^{(k)}$ and units is less straightforward to study, as is the final implication that a matrix that maps $\mathcal{O}_{\mathbb{K}}^{\times}$ to itself must be of the expected form. However, using the above described techniques, including the algebraic-geometric characterization of $W^{(k)}$, this result has been proved for all cyclotomic fields. Since the details are significantly more technical than in the quadratic case, they fall outside of the scope of this summary, so we will not delve into further detail.

Acknowledgements

MB would like to thank Valentin Blomer for helpful discussions and hints on the manuscript. We thank the referee for a number of suggestions, which helped to improve the presentation. Moreover, AB is grateful to ANID (formerly CONICYT) for the financial support received under the Doctoral Fellowship ANID-PFCHA/Doctorado Nacional/2017-21171061, ML was supported by Narodowe Centrum Nauki grant UMO-2019/33/B/ST1/00364, and AN acknowledges financial support by the German Research Foundation (DFG) through its Heisenberg program (project number 437113953).

[3] All quadratic fields are Galois, so we use the notation $\mathrm{Gal}(\mathbb{K}|\mathbb{Q})$ instead of $\mathrm{Aut}(\mathbb{K}|\mathbb{Q})$.

Part IV

Appendix

Conclusions

The extended symmetry group are a powerful and versatile algebraic invariant, whose structure may reveal the properties of the underlying shift space that cannot be "detected" by automorphisms, such as e.g. palindromicity or invariance under certain affine transformations. They take special prominence in the low-complexity situation, where other invariants such as entropy fail to distinguish between different systems, while also being a sort of "intermediate" step between the automorphism group and other more complicated algebraic objects defined in relation to a dynamical system, like full groups.

Over the course of the development of this thesis work, several connections between symbolic dynamics and other branches of mathematics have been found, starting from the obvious links to classical geometry and the theory of rigid transformations. From then on, we see connections to the theory of finite groups in the theory of automorphisms and extended symmetries of shifts generated by bijective substitutions; as we saw in Chapters 4 and 5, both $\mathrm{Aut}(\mathsf{X}_\theta, \mathbb{Z}^d)$ and $\mathrm{Sym}(\mathsf{X}_\theta, \mathbb{Z}^d)$ are linked to the column groups defined by the substitution, their centralizers and normalizers as subgroups of a permutation group.

In Chapter 6, we proceed further and define a class of subshifts defined via divisibility properties and other number-theoretical criteria. There, we found that these groups and the associated spaces are connected with the structure of the underlying ring, and that maps that preserve Dedekind factorizations in turn induce extended symmetries. Furthermore, the theory that leads to such a connection is linked to various key features of algebraic and analytic number theory, such as the properties of field norms and Galois automorphisms, the Dedekind zeta function, and the Minkowski embedding; moreso, for fields of higher degree, we even need to reach into the basics of algebraic geometry, as the field norm is analyzed in terms of the algebraic varieties defined by it, their decomposition into irreducible varieties, and their relation to the aforementioned Minkowski embedding.

Thus, we see that the theory of extended symmetries, even restricted to the scope of symbolic dynamics, is deep enough to serve as a "hub" for such kinds of connections to appear, and in turn, this makes symbolic dynamics a tool appropriate to approach problems from other areas of mathematics. Examples include the analysis of the half-hex tiling and other similar ones as seen in, e.g., Chapter 4 and Baake and Grimm's work on the squiral tiling [7]; another well-known example, into which we do not delve here but it is worth mentioning, is the proof of Furstenberg's recurrence theorem and its close relationship to the theory of arithmetic progressions and Szemerédi's theorem in number theory. We expect that this work may be taken as an initial step to introduce ideas from the theory of extended symmetries for these purposes, for example, in order to analyze the local structure of the set V of visible points via the associated shift space X_V.

We expect this line of work to allow for further research in several directions. For example, part of the theory of bijective substitutions seen in Chapters 4 and 5 ought to be extended to other kinds of substitutions, such as those with coincidences, that would require a different approach. We expect that the underlying quasi-periodic or Toeplitz structure that might appear in such shift spaces will play a role in this situation. Furthermore, one could take this approach as a template, in order to devise how to deal with general **Toeplitz shifts**, which are also hierarchical in nature. Other possible directions include different kinds of substitutions, such as those with overlap (see e.g. the Robinson tiling itself [43, 44]) or **digit tilings**, which are substitutions with non-rectangular support, linked with a linear **expansive matrix**, which is nondiagonal and of determinant $\Delta > 1$, which establishes how lower-order supertiles fit together to conform higher-order ones. The underlying theory of automorphisms in such examples appears to be fairly similar to the rectangular situation, but, as the limit shape of the supertiles is often fractal and may have gaps (and even be disconnected), we expect the geometry of such a shift space to be much richer, leading to more variety among the associated extended symmetry groups.

Similarly, the results on k-free shifts shown in Chapter 6 are already known to be generalizable to other classes of fields, most notably all quadratic and cyclotomic fields; the former may be done with elementary number-theoretical characterizations of the field norm, but it appears that algebraic geometry is a required tool for the study of the latter. The obtained description of the extended symmetry group follows the same form as the examples in Chapter 6, being entirely determined by the unit structure and Galois automorphisms of the underlying ring. This suggests that such a description may apply to all Galois extensions of the rationals, or even maybe all algebraic fields and their associated rings of integers. Other directions of study include defining analogous shift spaces using non-zero characteristic fields as a basis, or even other algebraic structures that might have lattice-based substructures with similar behavior to that of Dedekind factorizations in an integer ring, e.g. Gaussian semigroups. As well, the theory of $\mathcal{B}$-free shifts, even in the $\mathbb{Z}$ case, appears to be much richer when the Erdős condition on the corresponding lattices is removed, with more variety in the kind of automorphisms that may appear; this, of course, should lead to a similar increase of complexity in the case of extended symmetries in dimension 2 and higher.

It is also worth mentioning that most work on extended symmetries has focused on $\mathbb{Z}$ or $\mathbb{Z}^d$ as the underlying group. In the general setting of subshifts defined over a group G, the role of $\mathrm{GL}_d(\mathbb{Z})$ is overtaken by the set of group automorphisms $\mathrm{Aut}(G)$, and a similar theory of extended symmetries can be developed; however, not much is known in this direction. In the case where G is non-abelian, the relationship between shift maps and automorphisms is more complicated: not all shift maps σ_g are shift automorphisms, due to the lack of commutativity. However, all shift maps are extended symmetries. This hints at a deeper complexity in the relation between the two groups. In this direction, tools such as **projective subdynamics** might be good ways to establish the nature of such a connection, by reducing the problem to the analysis of the behavior of such maps in a subgroup of lower complexity.

In summary, extended symmetries are both a powerful tool to study shift spaces and a rich object of study in itself, and we expect that the research outlined during this thesis work can be continued in several different ways, both in the realm of “pure” symbolic dynamics and in connection to several other areas of mathematics.

Appendix A

Samples of exploratory code

This small section focuses on listing some of the mathematical scripts used during the development of this thesis work, including the implementation of a few algorithms discussed in Chapters 4 and 5. Several of the figures included in this thesis work were generated using these scripts as a base. They are developed in the `Sagemath`® mathematical environment, which is based on the `Python`® programming language.

Observation: Most of the code here is designed as a proof of concept and only tested for small cases, and thus efficiency was not a priority. This is kept as a sample of the capabilities of the `Sagemath`® environment as a tool for research.

A.1. One-dimensional substitutions

In this part, substitutions are represented as dictionaries, where keys are one-character strings (letters of an alphabet) and the associated values are the corresponding substituted words. For example, the Fibonacci substitution corresponds to:

```
Fib_sub = {"0" : "01",
           "1" : "0"}
```

A.1.1. Basic substitutive operations

The following code iterates a substitution over a string.

```
def str_subst(in_str,rule):
  lst_out = [rule[a] for a in list(in_str)]
  return "".join(lst_out)

def iterated_str_subst(in_str,rule,n):
  output = in_str;
  for k in range(0,n):
    output = str_subst(output,rule)
  return output
```

The next function computes the n-th power of a substitution.

```
def subst_power(rule_dict,power):
  return {key : iterated_str_subst(key,rule_dict,power)
              for key in rule_dict}
```

This is a function devised to obtain a substitutive word $\theta^k(a)$ containing a given word w; this was used as a subroutine for a (partially implemented and finally unused) procedure to find a desubstitution for a word.

```
def subst_get_least_word(word,
                         rule_dict,
                         alphabet,
                         constant_length=True,
                         tolerance=10):
  subst_words = copy(alphabet)
  if constant_length == True:
    tolerance = len(alphabet)^2 +
              ceil(log(len(word) /
              (subst_length_const(rule_dict) - 1))) + 1
  for i in range(tolerance):
    for sw in subst_words:
      if word in sw:
        return sw
      subst_words = [str_subst(w,rule_dict)
                   for w in subst_words]
  return ""
```

The following computes the substitution matrix associated to a substitution:

```
def subst_matrix(subst_rule,alphabet):
  M = [[0 for _ in alphabet] for _ in alphabet]
  N = len(alphabet)
  for i in range(N):
    for row in subst_rule[alphabet[i]]:
      for entry in row:
        M[i][alphabet.index(entry)] =
          M[i][alphabet.index(entry)] + 1
  return matrix(M)
```

A.1.2. Constant length substitutions

The following function verifies whether a substitution is of constant length and returns its length.

```
def subst_length_const(rule_dict):
  slen = -1
  for key in rule_dict:
    if(slen == -1):
      slen = len(rule_dict[key])
    else:
      if(len(rule_dict[key]) != slen):
        return -1
  return slen
```

The following code returns the columns of a (bijective) substitution, either as a dictionary or as a Permutation object.

```
def subst_column_dict(rule_dict, col):
  return {rulekey : rule_dict[rulekey][col]
                          for rulekey in rule_dict}


def subst_column_permutation(rule_dict, col):
  col_dict = subst_column_dict(rule_dict,col)
  alphabet = col_dict.keys()
  images = [col_dict[key] for key in alphabet]
  SA = SymmetricGroup(domain=alphabet)
  return SA(images)
```

The following reconstructs the substitution from its columns:

```
def subst_from_columns(col_list):
  return {a : "".join([col(a) for col in col_list])
              for a in col_list[0].parent().domain()}
```

The following verifies whether a substitution is bijective:

```
def subst_is_bijective(rule_dict):
  N = subst_length_const(rule_dict)
  if(N == -1):
    return False
  for i in range(N):
    col = subst_column_dict(rule_dict,i)
    vals = []
    for key in col:
      if col[key] not in vals:
        vals.append(col[key])
      else:
        return False
  return True
```

This computes the n-th column group of a substitution:

```
def subst_column_group(rule_dict, iterations):
  alphabet = rule_dict.keys()
  slen = subst_length_const(rule_dict)
  rule_pow = subst_power(rule_dict,iterations)
  return PermutationGroup(
    [subst_column_permutation(rule_pow, col)
     for col in range(slen^iterations)],
    domain=alphabet)
```

A.1.3. Automorphisms and extended symmetries of constant length substitutions

This is an implementation of Algorithm 1 in Chapter 5, computing the letter swaps associated to the automorphism group of a bijective substitution.

```
def subst_letter_swaps_group(rule_dict):
  alphabet = rule_dict.keys()
  slen = subst_length_const(rule_dict)
  N = min([subst_column_permutation(rule_dict,i).order()
           for i in range(slen)])
  rule_pow = subst_power(rule_dict,N)
  SA = SymmetricGroup(domain=alphabet)
  swap_group = SA.centralizer(
               subst_column_permutation(rule_pow,0))
  for j in range(1,slen^N):
    swap_group = swap_group.intersection(
                 SA.centralizer(
                 subst_column_permutation(rule_pow,j)))
  return SA.subgroup([SA([alphabet[tpl(k)-1]
                          for k in range(1,len(alphabet)+1)])
                      for tpl in swap_group.gens()])
```

The following is an implementation of Algorithm 2 in Chapter 5, computing alphabet permutations that correspond to reversing symmetries.

```
def perm_conj_matching(perm1,perm2):
  assert perm1.conjugacy_class() == perm2.conjugacy_class(),
         "Permutations have to be conjugate to be matched"
  symg = perm1.parent()
  dom = list(symg.domain())
  L1 = perm1.cycle_tuples()
  L2 = perm2.cycle_tuples()
  L1.sort(key = len)
  L2.sort(key = len)
  equiv_dict = {L1[i][j] : L2[i][j]
                    for i in range(len(L1))
                    for j in range(len(L1[i]))}
  return symg([equiv_dict[x] if x in equiv_dict else x
               for x in dom])

def perm_mirror_coset(perm1,perm2):
  assert perm1.conjugacy_class() == perm2.conjugacy_class(),
         "Permutations have to be conjugate to be matched"
  symg = perm1.parent()
  matching_map = symg(perm_conj_matching(perm1,perm2))
  coset_left = Set([symg(c)*matching_map
                    for c in symg.centralizer(perm1)])
  return coset_left
```

```
def mirror_permutations(subst_rule):
  N = subst_length_const(subst_rule)
  M = min([lcm(subst_column_permutation(subst_rule,i).order(),
           subst_column_permutation(subst_rule,N-i-1).order())
           for i in range(N/2)])
  rpow = subst_power(subst_rule,M)
  cols = [subst_column_permutation(rpow,i) for i in range(N^M)]
  coset_intersection = perm_mirror_coset(cols[0],cols[N^M-1])
  for i in range(1,N^M/2):
    coset_inter = coset_inter.intersection(
                  perm_mirror_coset(cols[i],cols[N^M-i-1]))
  return coset_inter
```

A.2. Two-dimensional substitutions

Here, two-dimensional rectangular configurations are treated as matrices with integer values, and thus alphabets are always sets of integers. Thus, a substitution is represented as just a list or tuple of matrices, e.g. the symbolic half-hex representation is associated to the following substitution:

```
halfhex_sym = [matrix([[7,12],[5,0]]),
               matrix([[3,9],[1,4]]),
               matrix([[3,10],[1,4]]),
               matrix([[7,12],[5,3]]),
               matrix([[4,8],[0,2]]),
               matrix([[5,8],[0,2]]),
               matrix([[3,9],[6,4]]),
               matrix([[3,11],[6,4]]),
               matrix([[7,12],[5,8]]),
               matrix([[3,9],[9,4]]),
               matrix([[3,10],[9,4]]),
               matrix([[3,11],[9,4]]),
               matrix([[12,8],[0,2]])]
```

A.2.1. Basic substitutive operations

The following code iterates a substitution over a set of matrices.

```
def matrix_subst(M,dict_subst):
  M_lst = M.list()
  m,n = M.dimensions()
  entries = [dict_subst[k] for k in M_lst]
  return block_matrix(m,n,entries,subdivide=False)

def iterated_matrix_subst(M,dict_subst,n):
  P = M
  for _ in range(0,n):
    P = matrix_subst(P,dict_subst)
  return P
```

A.2.2. Analysis of structure substitutions

The following procedure marks a periodic subconfiguration on a rectangular pattern, that can come from a Toeplitz structure (see e.g. substitutions with coincidences, or the half-hex rectangular representation):

```
def largest_periodic_subset(mat_tiling,period_matrix):
  mx,my = mat_tiling.dimensions()
  period_SNF = period_matrix.smith_form()
  smcoeff = period_SNF[0].diagonal()
  per_candidates = [[-1 for j in range(smcoeff[1])]
                    for i in range(smcoeff[0])]
  can_be_periodic = [[True for j in range(smcoeff[1])]
                     for i in range(smcoeff[0])]
  for i in range(mx):
    for j in range(my):
      imd, jmd = congruence_class_mod_matrix_smith(
                   vector([i,j]),
                   period_SNF)
      if not can_be_periodic[imd][jmd]:
        continue
      if per_candidates[imd][jmd] == -1:
        per_candidates[imd][jmd] = mat_tiling[i,j]
      else:
        if per_candidates[imd][jmd] != mat_tiling[i,j]:
          can_be_periodic[imd][jmd] = False
  matrix_perset = matrix([[-1 for j in range(my)]
                          for i in range(mx)])
  for i in range(mx):
    for j in range(my):
      imd, jmd = congruence_class_mod_matrix_smith(
                   vector([i,j]),
                   period_SNF)
      matrix_perset[i,j] = Integer(can_be_periodic[imd][jmd])
  return matrix_perset
```

The next function is designed to help compute the height lattice of a substitution, by checking whether the alphabet of a matrix substitution is partitioned along cosets of a given lattice.

```
def lattice_alph_part(tiling,lat_mat):
  tx, ty = tiling.dimensions()
  dict_eqclass = {}
  SNF = lat_mat.smith_form()
  for i in range(tx):
    for j in range(ty):
      key = tuple(congruence_class_mod_matrix_smith(
                    vector([i,j]),
                    SNF))
      if key in dict_eqclass:
        dict_eqclass[key].add(tiling[i,j])
      else:
        dict_eqclass[key] = set([tiling[i,j]])
      return dict_eqclass
```

A.3. Tilings

Tilings are represented as lists of ordered pairs $(\vec{x}, n)$, where $\vec{x}$ is the position of a tile of type n (with n an integer). Thus, inflation rules are lists of tilings, together with an inflation factor. For instance, the following code constructs the representation of the half-hex inflation as a list of 6 tilings:

```
hex_vertices = [vector([cos(k*pi/3), sin(k*pi/3)])
                for k in range(6)]

rule = [[[[0,0],i]] +
       [[list(hex_vertices[mod(i+j-1,6)] +
              hex_vertices[mod(i+j,6)]),
         mod(i+j+1,6)]
         for j in range(1,4)] for i in range(6)]

inflation = 2
```

The following implements inflation rules for tilings:

```
def inflation_rule(tiling,rule,inflation_factor):
  inflated_tiling = [[[tile[0][0]*inflation_factor,
                       tile[0][1]*inflation_factor],
                       tile[1]] for tile in tiling]
  new_tiling = []
  for tile in inflated_tiling:
    newtiles = [[[x+y for x,y in zip(tile[0],rep_tile[0])],
                 rep_tile[1]]
                 for rep_tile in rule[tile[1]]]
    new_tiling = new_tiling + newtiles
  return new_tiling
```

The following function plots a tiling:

```
def tiling_plot(tiling,shapes,colors,edgecol):
  G = plot([],axes=False)
  for tile in tiling:
    G = G + polygon([[x+y for x,y in zip(tile[0],shape_point)]
                     for shape_point in shapes[tile[1]]],
                     rgbcolor=colors[tile[1]],
                     edgecolor=edgecol,
                     thickness=1)
  return G
```

A.4. Algebraic shift spaces

The following is an example on how to use Sage to find a list of k-free numbers in an algebraic integer ring such as $\mathbb{Z}[\frac{1}{2}(1+\sqrt{5})]$.

```
def is_ideally_kfree(field,intring_number,exponent):
  if intring_number == 0:
    return False
  if abs(intring_number.norm()) == 1:
    return True
  for factor_pair in list(
    field.fractional_ideal(intring_number).factor()):
    if abs(factor_pair[0].norm()) == 1:
      continue
    if factor_pair[1] >= exponent:
      return False
    return True

UK = NumberField(x^2 + 5, ’u’)
uu = UK(’u’)

N = 10
M = matrix([[int(is_ideally_kfree(UK,b*uu + a,2))
             for a in range(-N,N+1)]
             for b in range(-N,N+1)])
```

Bibliography

[1] T. M. Apostol. *Introduction to Analytic Number Theory.* Springer, New York, 1984.

[2] M. Baake. Structure and representations of the hyperoctahedral group. *J. Math. Phys.*, 25:3171–3184, 1984.

[3] M. Baake. A brief guide to reversing and extended symmetries of dynamical systems. In S. Ferenczi, J. Kulaga-Przymus, and M. Lemańczyk, editors, *Ergodic Theory and Dynamical Systems in their Interactions with Arithmetics and Combinatorics*, pages 117–135. Springer, Cham, 2018.

[4] M. Baake, A. Bustos, C. Huck, M. Lemańczyk, and A. Nickel. Number-theoretic positive entropy shifts with small centraliser and large normaliser. *Ergodic Theory Dynam. Systems*, 2019.

[5] M. Baake, Á. Bustos, and A. Nickel. On the stabiliser of some number-theoretic shift spaces. in preparation.

[6] M. Baake and U. Grimm. *Aperiodic Order, vol. 1: A Mathematical Invitation.* Cambridge University Press, Cambridge, 2013.

[7] M. Baake and U. Grimm. Squirals and beyond: substitution tilings with singular continuous spectrum. *Ergodic Theory Dynam. Systems*, 34(4):1077–1102, 2014.

[8] M. Baake and C. Huck. Ergodic properties of visible lattice points. *Proc. V. A. Steklov Inst. Math.*, 288:184–208, 2015.

[9] M. Baake, C. Huck, and N. Strungaru. On weak model sets of extremal density. *Indag. Math.*, 28:3–31, 2017.

[10] M. Baake and D. Lenz. Dynamical systems on translation bounded measure: pure point dynamical and diffraction spectra,. *Ergodic Theory Dynam. Systems*, 24:1867–1893, 2004.

[11] M. Baake, R. V. Moody, and P. A. B. Pleasants. Diffraction of visible lattice points and k-th power free integers. *Discr. Math.*, 221:3–42, 2000.

[12] M. Baake and J. A. G. Roberts. The structure of reversing symmetry groups. *Bull. Austral. Math. Soc.*, 73:445–459, 2006.

[13] M. Baake, J. A. G. Roberts, and R. Yassawi. Reversing and extended symmetries of shift spaces. *Discrete Contin. Dyn. Syst.*, 38(2):835–866, 2018.

[14] M. Baake, T. Spindeler, and N. Strungaru. Diffraction of compatible random substitutions in one dimension. *Indag. Math.*, 29:1031–1071, 2018.

[15] S. Bezuglyi and K. Medynets. Full groups, flip conjugacy, and orbit equivalence of cantor minimal systems. *Colloq. Math.*, 110:409–429, 2008.

[16] V. Blomer. private communication, 2017.

[17] Z. I. Borevich and I. R. Shafarevich. *Number Theory.* Academic Press, New York, 1966.

[18] M. Boyle. *Topological orbit equivalence and factor maps in symbolic dynamics.* PhD thesis, University of Washington, 1983.

[19] M. Boyle, D. Lind, and D. Rudolph. The automorphism group of a shift of finite type. *Trans. Amer. Math. Soc.*, 306(1):71–114, 1988.

[20] M. Boyle and J. Tomiyama. Bounded topological orbit equivalence and C^*-algebras. *J. Math. Soc. Japan*, 50(2):317–329, 1998.

[21] M. Burrow. *Representation theory of finite groups.* Dover, New York, 1993.

[22] A. Bustos. Extended symmetry groups for multidimensional subshifts with hierarchical structure. *Discr. Cont. Dynam. Syst. A*, 40:5869–5895, 2020.

[23] Á. Bustos, D. Luz, and N. Mañibo. Admissible reversing and extended symmetries for bijective substitutions. *arXiv e-prints*, 2021.

[24] T. Ceccherini-Silberstein and M. Coornaert. *Cellular Automata and Groups.* Springer, Berlin, 2010.

[25] F. Cellarosi and I. Vinogradov. Ergodic properties of k-free integers in number fields. *J. Mod. Dyn.*, 7:461–488, 2013.

[26] M. I. Cortez and F. Durand. Self-similar tiling systems, topological factors and stretching factors. *Discr. Comput. Geom.*, 40:622–640, 2008.

[27] M. I. Cortez and S. Petite. Realization of big centralizers of minimal aperiodic actions on the cantor set. *Discr. Cont. Dynam. Syst. A*, 40:2891–2901, 2020.

[28] E. M. Coven. Endomorphisms of substitution minimal sets. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 20(2):129–133, 1971.

[29] E. M. Coven, A. Quas, and R. Yassawi. Computing automorphism groups of shifts using atypical equivalence classes. *Discrete Anal.*, 3, 2016.

[30] V. Cyr and B. Kra. The automorphism group of a shift of linear growth: beyond transitivity. *Forum Math. Sigma*, 3:e5, 2015.

[31] V. Cyr and B. Kra. The automorphism group of a shift of subquadratic growth. *Proc. Amer. Math. Soc.*, 144:613–621, 2016.

[32] S. de Neymet. *Introducción a los grupos topológicos de transformaciones.* Sociedad Matemática Mexicana, Ciudad de México, 2005.

[33] S. Donoso, F. Durand, A. Maass, and S. Petite. On automorphism groups of low complexity subshifts. *Ergodic Theory Dynam. Systems*, 36(1):64–95, 2015.

[34] S. Donoso and W. Sun. Dynamical cubes and a criteria for systems having product extensions. *J. Mod. Dyn.*, 9:365–405, 2015.

[35] F. Durand and J. Leroy. Decidability of the isomorphism and the factorization between minimal substitution subshifts. *arXiv e-prints*, 2018.

[36] A. Dymek, Kasjan S., Kułaga Przymus J., and Lemańczyk M. $\mathcal{B}$-free sets and dynamics. *Trans. Amer. Math. Soc.*, 370:5425–5489, 2018.

[37] E. H. El Abdalaoui, M. Lemańczyk, and T. de la Rue. A dynamical point of view on the set of $\mathcal{B}$-free integers. *Intern. Math. Res. Notices*, 16:7258–7286, 2015.

[38] N. P. Fogg, editor. *Substitutions in Dynamics, Arithmetics and Combinatorics.* Springer, Berlin, 2002.

[39] R. Fokkink and R. Yassawi. Topological rigidity of linear cellular automaton shifts. *Indag. Math.*, 29:1105–1113, 2018.

[40] N. P. Frank. Substitution sequences in $\mathbb{Z}^d$ with a non-simple lebesgue component in the spectrum. *Ergodic Theory Dynam. Systems*, 23:519–532, 2003.

[41] N. P. Frank. Multidimensional constant-length substitution sequences. *Topology Appl.*, 152(1):44–69, 2005.

[42] N. P. Frank. Introduction to hierarchical tiling dynamical systems. In S. Akiyama and P. Arnoux, editors, *Substitution and Tiling Dynamics: Introduction to Self-inducing Structures*, pages 33–95. Springer, 2020.

[43] F. Gähler. Substitution rules and topological properties of the Robinson tilings. In Siegbert Schmid, Ray L. Withers, and Ron Lifshitz, editors, *Aperiodic Crystals*, pages 67–73. Springer, Netherlands, 2013.

[44] F. Gähler, A. Julien, and J. Savinien. Combinatorics and topology of the Robinson tiling. *C. R. Math. Acad. Sci. Paris*, 350(11):627–631, 2012.

[45] A. Gathmann. *Algebraic Geometry.* Technische Universität Kaiserslautern, Kaiserslautern, 2020.

[46] G. Goodson, A. del Junco, M. Lemańczyk, and D. Rudolph. Ergodic transformations conjugate to their inverses by involutions. *Ergodic Theory Dynam. Systems*, 16:97–124, 1996.

[47] G. R. Goodson. Inverse conjugacies and reversing symmetry groups. *Amer. Math. Monthly*, 106(1):19–26, 1999.

[48] P. A. Grillet. *Abstract Algebra.* Springer, New York, 2007.

[49] M. Hall. *The Theory of Groups.* Chelsea Publishing Company, New York, 1976.

[50] G. M. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers.* Oxford University Press, Oxford, 2008.

[51] R. Hartshorne. *Algebraic Geometry.* Springer, New York, 2010.

[52] C. Holton, C. Radin, and L. Sadun. Conjugacies for tiling dynamical systems. *Commun. Math. Phys.*, 254:343–359, 2005.

[53] T. Hungerford. *Algebra.* Springer, New York, 2003.

[54] G. James and A. Kerber. *The Representation Theory of the Symmetric Group.* Addison-Wesley Publishing Company, Reading, Massachusetts, 1981.

[55] F. Jarvis. *Algebraic Number Theory.* Springer, Switzerland, 2014.

[56] S. Kasjan, G. Keller, and M. Lemańczyk. Dynamics of $\mathcal{B}$-free sets: a view through the window. *Intern. Math. Res. Notices*, 9:2690–2734, 2019.

[57] J. Kellendonk and R. Yassawi. The ellis semigroup of bijective substitutions. *arXiv e-prints*, 2019.

[58] G. Keller. private communication, 2019.

[59] G. Keller. Maximal equicontinuous generic factors and weak model sets. *Discr. Cont. Dynam. Syst. A*, 40:6855–6875, 2020.

[60] K. Kendig. *Elementary Algebraic Geometry.* Dover, New York, 2015.

[61] Y.-O. Kim, J. Lee, and K. K. Park. A zeta function for flip systems. *Pacific J. Math*, 209:289–301, 2003.

[62] B. Kitchens. *Symbolic dynamics: one-sided, two-sided and countable state Markov shifts.* Springer, Berlin, 1998.

[63] B. Kitchens and K. Schmidt. Isomorphism rigidity of simple algebraic $\mathbb{Z}^d$-actions. *Invent. Math.*, 142:559–577, 2000.

[64] P. Kůrka. *Topological and Symbolic Dynamics.* Société mathématique de France, Paris, 2003.

[65] S. Labbé. Spectral theory of $\mathbb{Z}^d$ substitutions. *Ergodic Theory Dynam. Systems*, 38:1289–1341, 2018.

[66] S. Labbé. Substitutive structure of jeandel?rao aperiodic tilings. *Discrete Comput. Geom.*, 2019.

[67] J.S.W. Lamb and J.A.G. Roberts. Time-reversal symmetry in dynamical systems: a survey. *Physica D*, 112:1–39, 1998.

[68] S. Lang. *Algebra.* Springer, New York, 2002.

[69] S. Lang. *Introduction to Algebraic Geometry.* Dover, New York, 2019.

[70] M. Lemańczyk and M. K. Mentzen. On metric properties of substitutions. *Compositio Math*, 65(3):241–263, 1988.

[71] D. Lind and B. Marcus. *An Introduction to Symbolic Dynamics and Coding.* Cambridge University Press, New York, 1995.

[72] M. Lothaire, editor. *Combinatorics on Words.* Cambridge University Press, Cambridge, 1997.

[73] M. Lothaire, editor. *Algebraic Combinatorics on Words.* Cambridge University Press, Cambridge, 2002.

[74] S. Mac Lane. *Categories for the Working Mathematician.* Springer, New York, 1978.

[75] G. Maloney and D. Rust. Beyond primitivity for one-dimensional substitution subshifts and tiling spaces. *Ergodic Theory Dynam. Systems*, 38(3):1086–1117, 2018.

[76] F. B. Martinez, C. G. Moreira, N. Saldanha, and E. Tengan. *Teoria dos números: um passeio com primos e outros números familiares pelo mundo inteiro.* Instituto Nacional de Matemática Pura e Aplicada, Rio de Janeiro, 2013.

[77] M. K. Mentzen. Automorphisms of subshifts defined by $\mathcal{B}$-free sets of integers. *Coll. Math.*, 147:87–94, 2017.

[78] G. A. Miller. Groups formed by special matrices. *Bull. Am. Math. Soc.*, 24:203–206, 1918.

[79] B. Mossé. Reconnaissabilité des substitutions et complexité des suites automatiques. *Bull. Soc. Math. France*, 124(2):329–346, 1996.

[80] C. Müllner and R. Yassawi. Automorphisms of automatic shifts. *Ergodic Theory Dynam. Systems*, 41:1530–1559, 2021.

[81] J. Neukirch. *Algebraic Number Theory.* Springer, Berlin, 1999.

[82] A.G. O'Farrell and I. Short. *Reversibility in Dynamics and Group Theory.* Cambridge University Press, Cambridge, 2015.

[83] J. Olli. Endomorphisms of Sturmian systems and the discrete chair substitution tiling system. *Discrete Cont. Dyn. Syst.*, 33(9):4173–4186, 2013.

[84] P. A. B. Pleasants and C. Huck. Entropy and diffraction of the k-free points in n-dimensional lattices. *Discr. Comput. Geom.*, 50:39–68, 2013.

[85] M. Queffélec. *Substitution Dynamical Systems–Spectral Analysis.* Springer, Berlin, 2010.

[86] E. Riehl. *Category Theory in Context.* Dover, New York, 2016.

[87] J. A. G. Roberts and G. R. W. Quispel. Chaos and time-reversal symmetry: order and chaos in reversible dynamical systems. *Phys. Rep.*, 216:63–177, 1992.

[88] R. M. Robinson. Undecidability and nonperiodicity for tilings of the plane. *Invent. Math.*, 12:177–209, 1971.

[89] L. Sadun. *Topology of Tiling Spaces.* American Mathematical Society, Rhode Island, 2008.

[90] K. Schmidt. *Dynamical Systems of Algebraic Origin.* Birkhäuser, Basel, 1995.

[91] I. Segal. The automorphisms of the symmetric group. *Bull. Amer. Math. Soc.*, 46:565, 1940.

[92] S. K. Sehgal. On the normalizer of a group in the cayley representation. *Int. J. Math. Math. Sci.*, 12:459–462, 1989.

[93] I. R. Shafarevich. *Basic Algebraic Geometry 1: Varieties in Projective Space.* Springer, Berlin, 2013.

[94] B. Solomyak. Nonperiodicity implies unique composition for self-similar translationally finite tilings. *Discrete Comput. Geom.*, 20(2):265–279, 1998.

[95] I. Vaisencher. *Introdução as Curvas Algébricas Planas.* Instituto Nacional de Matemática Pura e Aplicada, Rio de Janeiro, 2009.

[96] D. Zagier. *Zetafunktionen und quadratische Körper.* Springer, Berlin, 1981.

Index