Universität Leipzig
Fakultät für Mathematik und Informatik
Institut für Informatik

A generalization of the iteration theorem for recognizable formal power series on trees

Bachelorarbeit

Leipzig, 24. September 2019

vorgelegt von
Patrick Kramer
im Studiengang B.Sc. Informatik

Betreuer: Prof. Dr. Andreas Maletti
Abteilung: Algebraische und logische Grundlagen

Contents

Abstract

Berstel and Reutenauer stated the iteration theorem for recognizable formal power series on trees over fields and vector spaces [BR80]. The key idea of its proof is the existence of pseudo-regular matrices in matrix-products [Jac80]. This theorem is generalized to integral domains and modules over integral domains in this thesis. It only requires the reader to have basic knowledge in linear algebra. Concepts from the advanced linear algebra and abstract algebra are introduced in the preliminary chapter.

1 Introduction

Trees over an alphabet appear as natural generalization of common words over an alphabet. Indeed, many concepts known from formal languages on words can be transferred to tree languages. Formal power series are a well-known concept in the field of study of formal languages. A formal power series is a function that maps every element of the freely generated monoid over an alphabet to an element taken from a field [SS78]. Hence, the name *word function* would also be suitable to denote formal power series. As somebody could guess, there is a generalization to trees, namely, *formal power series on trees* or *tree functions* that is a function mapping each tree to an element taken from a field. Tree functions can be used to do calculations on trees, e.g. we could calculate the height of a tree or evaluate arithmetic expressions [BR80]. There is a classification of formal power series depending on how easily function values of this tree function can be computed. If we are able to calculate function values by multilinear functions over a vector space, that is, the formal power series is easy to compute, we call it *recognizable*. Moreover, an arbitrary formal power series can generate a language over trees, namely, the set containing all trees that are not mapped to zero, which is called the *support* of the formal power series.

There is an iteration theorem for word functions. It states that we can take any word w in the support of a recognizable word function and then there is a part v in the word $w = uvw$ such that uv^kw is an element of the support for infinitely many k. Gérard Jacob proved this over finite alphabets first [Jac80]. Reutenauer made Jacob's proof shorter and additionally presented a proof that generalizes the iteration theorem to infinite alphabets [Reu80]. This iteration theorem also holds for formal power series on trees. It again states that we can take any tree t in the support of a recognizable tree function and repeat a specific part of a long-enough walk in the tree such that an infinite number of such iterations lies in the support. The proof is based on the same idea as on words. Indeed, Berstel and Reutenauer proved it by using the statements about pseudo-regular matrices [BR80].

The aim of this thesis is to generalize Berstel's and Reutenauer's iteration theorem from fields to integral domains. Indeed a formal power series now maps the trees to elements taken from an integral domain. Also the notion *recognizable* is now defined by modules over integral domains instead of vector spaces. Therefore we have a weaker underlying weight-structure on our tree functions, but the iteration theorem still holds. For instance, we could define a formal power series that maps to the integers $\mathbb{Z}$, which is an integral domain, and we can still apply the iteration

theorem. By generalizing the result to integral domains, we also obtain the assertion over integrally closed domains, GCD domains, unique factorization domains, principal ideal domains, euclidean domains and of course over fields. The reason for this is the following inclusion chain:

commutative rings
$\cup\!\!\!\!\neq$
integral domains
$\cup\!\!\!\!\neq$
integrally closed domains
$\cup\!\!\!\!\neq$
GCD domains
$\cup\!\!\!\!\neq$
unique factorization domains
$\cup\!\!\!\!\neq$
principal ideal domains
$\cup\!\!\!\!\neq$
euclidean domains
$\cup\!\!\!\!\neq$
fields.

Given a formal power series on trees over an integral domain, which is recognizable by a given module, the main idea of the generalization is to construct the field of fractions of the integral domain and apply the iteration theorem over fields on it. By showing that important statements about pseudo-regular matrices still hold for a matrix if the corresponding matrix over the field of fractions is pseudo-regular, we obtain the desired generalization.

A theoretical application of the iteration theorem is that we can use it to prove that some formal power series are not recognizable. For instance, arithmetic expressions with division included are not recognizable. Note that if we exclude the division it is recognizable [BR80]. There are also practical implications if we apply the iteration theorem on some specific formal power series: So if an arithmetic expression (that does not contain division) is long enough and is unequal to zero, we know that we can repeat a part in it such that it is still unequal to zero.

We begin this thesis with some preliminaries. These are interesting if the reader is not familiar with the basic and some advanced concepts of (linear) algebra. At first we introduce some fundamental algebraic structures that will be needed in this thesis. In particular, we discuss integral domains and modules over them. A main statement is that every free module can be assigned a well-defined dimension as known from vector spaces. Secondly, we introduce the field of fractions of an integral domain, which is the key aspect of the iteration theorem's generalization. The last preliminary section is about fundamental notions and statements about

matrices and homomorphisms: We present the concept of transformation matrices and minimal polynomials. Both will be necessary to define pseudo-regular matrices.

The next chapter will consider long products of matrices. After we define pseudo-regular matrices over fields, we show that every matrix-product that is long enough contains a pseudo-regular factor. This is not easy to prove since we need the tensor product and the exterior product, which indeed will be also introduced in this chapter. At the end of the chapter, we generalize this theorem to arbitrary integral domains. Also we prove a lemma, which states that a specific concatenation of pseudo-regular endomorphisms is unequal to zero if some conditions are satisfied. Indeed this is proved in the general case, that is, over endomorphisms on modules over integral domains. This lemma will be used to prove the generalized iteration theorem. In particular it helps to verify that the formal power series applied on the iterated tree is still unequal to zero for infinitely many iterations, hence these trees lie in the support of the formal power series.

After that, we are going to introduce formal power series on trees. Hence, Chapter 4 contains many definitions of basic notions and examples concerning tree functions.

Chapter 5 is the main chapter, which contains the generalized iteration theorem for recognizable formal power series on trees. Indeed, we prove this directly in the generalized version, that is, over integral domains and modules over them. We also consider an example.

The final chapter's aim is to summarize this thesis and to give ideas concerning future work on this field of study.

2 Preliminaries

We denote the set of natural numbers $\{0, 1, 2, ...\}$ by $\mathbb{N}$, whereby the zero is included. If we want to consider the natural numbers without zero we write down $\mathbb{N}_+ := \{1, 2, 3, ...\}$.

Let S be a finite set. We mean by $\#S$ the number of elements contained in S and call it the *cardinality* of S. If S is an infinite set we write down $\#S = \infty$. There is only one set that has cardinality zero, namely, the empty set, which we denote by $\emptyset$. We use the subset-symbols $\subseteq$ and $\subsetneq$ in the following way: $S \subseteq S'$ means that every element in S is also member of the set S'. $S \subsetneq S'$ means $S \subseteq S'$ and $S \neq S'$. $\mathcal{P}(M) := \{A \mid A \subseteq M\}$ denotes the powerset of M.

Let $n \in \mathbb{N}$ and $S_1, ..., S_n$ be non-empty sets. Then $S_1 \times ... \times S_n := \{(s_1, ..., s_n) \mid s_1 \in S_1, ..., s_n \in S_n\}$ defines the *Cartesian product* of $S_1, ..., S_n$. The members of this set are called *tuples*. If $S := S_1 = ... = S_n$ we also denote the Cartesian product by S^n. If A and B are two arbitrary non-empty sets, every $R \subseteq A \times B$ is called a *(binary) relation* on A and B. So if $a \in A$ and $b \in B$ are contained in the relation, that is, $(a, b) \in R$, we also use the infix notation and write down aRb.

Let S be an arbitrary set and $\leq$ be a binary relation on S^2. Then we call $(S, \leq)$ *partially ordered* if $\leq$ is reflexive, anti-symmetric and transitive. If $\leq$ is additionally connex, then we say that $(S, \leq)$ is *totally ordered*. A subset $C \subseteq S$ is called *chain* if $(C, \leq)$ is totally ordered. Indeed, we are able to compare each two elements taken from a chain. Moreover we call $(M, \leq)$ *inductively ordered* if every chain $C \subseteq S$ has an upper bound, and that means there exists $b \in S$ such that $c \leq b$ for all $c \in C$.

Given an arbitrary set M, we denote by Id_M the identity mapping on M, that is, $M \ni x \mapsto x \in M$.

We require Zermelo-Fraenkel set theory with the axiom of choice included (ZFC). There are many equivalent formulations of the axiom of choice, one is Zorn's lemma. It states that every non-empty inductively ordered set has a maximal element. We will leave out the proof that verifies the equivalence and just assume that Zorn's lemma holds.

2.1 Algebraic structures

The following introduction to algebraic structures is taken from Bosch's algebra book [Bos13] and the algebra book of Karpfinger and Meyberg [KM17]. The definitions for monoids and groups are written down for the sake of completeness. Fundamental statements about them, like the uniqueness of the identity element and of the inverse elements, are left out and should be known.

This section's main goal is to introduce integral domains and modules over them. Moreover, we want to prove basic statements, which we will need for the generalization of the iteration theorem for recognizable formal power series on trees to integral domains.

2.1.1 Monoids and groups

A monoid is a set together with an associative binary operation. The set is closed under this operation and contains an identity element.

Definition 2.1 (monoid)**.** We call $(X, *)$ a *monoid*, whereby X is a set and $* : X \times X \to X$ a mapping, if

- $*$ is associative, that is, $a * (b * c) = (a * b) * c$ for all $a, b, c \in X$ and
- there exists a so-called *identity element* $e \in X$ such that $x * e = e * x = x$ for all $x \in X$.

So a monoid does not include any properties about the existence of inverse elements. This motivates the definition of groups.

Definition 2.2 (group)**.** We call $(X, *)$ a *group* if $(X, *)$ is a monoid and every element has an *inverse element*, that is,

$$\forall x \in X \exists \tilde{x} \in X : x * \tilde{x} = e,$$

whereby e is the identity element of the monoid. If the multiplication commutes, we call the group *abelian*.

Sometimes we just want to write G instead of $(G, *)$. Then the operation symbol should be clear from the context. It can also happen that we use no symbol to denote the operation. Two notations of groups will be most important in the following, namely, the additive notation and the multiplicative notation. If we write a group in additive notation, the operation is denoted by $+$, the identity element by 0 and the inverse element of a by $-a$. Then we also write down $a - b$ instead of $a + (-b)$. This notation should be already known and therefore we leave out basic rules for it. In multiplicative notation we denote the operation by $\cdot$, the identity element by 1 and the inverse element of a by a^{-1}.

Remark. Actually it would be enough to postulate the existence of an element e that satisfies either $a * e = a$ or $e * a = a$ in the definition of groups. The corresponding other equation is an implication, even if we do not require the group to be abelian. But we want to keep it simple and just define groups with slightly stronger axioms.

Now we want to consider *factor groups*. We will need them in this chapter later on. The idea is to take a subgroup and to divide the group by this subgroup. What division exactly means in the case of groups will be clear by the next definitions.

We take the formal approach via equivalence classes to introduce factor groups. This has the advantage that we can use well-known results about equivalence relations. Indeed let G be a group and H be a subgroup, then we define $R \subseteq G \times G$ such that aRb if and only if $a - b \in H$. This relation is symmetric since $x \in H \Rightarrow -x \in H$, is reflexive since a subgroup contains the group's zero and transitive since aRb and bRc implies $a - c = (a - b) + (b - c) \in H$. We use the common notation $[a]_R$ to denote the equivalence class of a that is the set containing all elements of G that are in R-relation with a.

Definition 2.3 (factor group)**.** Let G be a group and H be a subgroup of G. Then the *factor group* G/H is defined as

$$G/H := \{[a]_R \mid a \in G\}.$$

Obviously we could also write $a + H := \{a + h \mid h \in H\}$ instead of $[a]_R$ since the sets are equal. Indeed every element b taken from $a + H$ satisfies aRb. Vice versa every element c taken from $[a]_R$ satisfies cRa, so $c - a \in H$, which means that there exists $h \in H$ such that $c = a + h$.

A well-known result for equivalence classes is that if $b \in [a]_R$, then $[a]_R = [b]_R$. In fact this means, for our relation R, that it is enough to show $a - b \in H$ to prove $a + H = b + H$. This also leads to the assertion that two equivalence classes are either equal or disjoint.

It is still not clear that the factor group actually forms a group. But indeed this is the case.

Lemma 2.1. *Let G be a group and H be a subgroup of G. Then $(G/H, \oplus)$ is a group, whereby*

$$(a + H) \oplus (b + H) := (a + b) + H.$$

If G is abelian, the factor group is also abelian.

Proof. $\oplus$ is well-defined because $a + H = a' + H$ and $b + H = b' + H$ implies $(a + b) - (a' + b') = (a - a') + (b - b') \in H$ which means $(a + b) + H = (a' + b') + H$. That G/H is closed under $\oplus$ is clear. Moreover, the associativity is transferred from G. Any element $a + H \in G/H$ has the inverse element $(-a) + H$ since $0 + H = H$ is the identity element of G/H.

If G is abelian, the commutativity is directly transferred to the factor group, hence it is also abelian. □

We used the notation $\oplus$ for the addition on the factor group to avoid coincidence with the addition $+$ over the group G. The difference should be clear now and we will also use the symbol $+$ instead of $\oplus$ from now on. But always remember the fact that these operations are still different. Let us consider two examples that will make the concept of the factor group clear.

Example. *We know that $(\mathbb{Z}, +)$ is an abelian group and $2\mathbb{Z} = \{2z \mid z \in \mathbb{Z}\}$ is a subgroup of it. Then*

$$\mathbb{Z}/2\mathbb{Z} = \{z + 2\mathbb{Z} \mid z \in \mathbb{Z}\} = \{2\mathbb{Z}, 2\mathbb{Z} + 1\} = \{\{0, 2, -2, 4, ...\}, \{1, -1, 3, ...\}\}.$$

The neutral element is $2\mathbb{Z}$. Both elements of the set are self-inverse.

Example. *The rational numbers with addition are an abelian group with the integers as a subgroup. If we consider the set $q + \mathbb{Z}$ for all $q \in \mathbb{Q}$, we notice that two such sets $q + \mathbb{Z}$ and $q' + \mathbb{Z}$ are equal if the digits after the point of q and q' are the same. Hence*

$$\mathbb{Q}/\mathbb{Z} = \{q + \mathbb{Z} \mid q \in [0, 1) \cap \mathbb{Q}\}.$$

The last definition concerning groups are structure preserving mappings over groups, the so-called *(group-)homomorphisms.*

Definition 2.4. Let $(G, *_G), (H, *_H)$ be groups and $\phi : G \to H$ be a function such that

$$\phi(x *_G x') = \phi(x) *_H \phi(x')$$

for all $x, x' \in G$. Then we call ϕ *(group-)homomorphism*. In this case:

- If ϕ is bijective, we also call ϕ *(group-)isomorphism* and G and H are called *isomorphic*.
- If $G = H$, we also call ϕ *(group-)endomorphism*. If the function is additionally bijective, we say *(group-)automorphism* to ϕ.
- If ϕ is injective, we also call ϕ *(group-)monomorphism*.
- If ϕ is surjective, we also call ϕ *(group-)epimorphism*.

We define the **kernel** $\operatorname{Ker}\phi$ of such a group-homomorphism ϕ by $\operatorname{Ker}\phi := \{x \in G \mid \phi(x) = e_H\}$, whereby e_H is the identity element of H.

2.1.2 Commutative rings and integral domains

If we want to define more than one operation over a set, the concept of *rings* gets important.

Definition 2.5 (commutative ring)**.** We call $(R, +, \cdot)$ a *commutative ring* if

- $(R, +)$ is an abelian group,
- $(R, \cdot)$ is a monoid,
- $0 \neq 1$,
- the multiplication commutes and
- multiplication distributes over addition, that is, $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ for all $a, b, c \in R$.

We call $+$ and $\cdot$ addition and multiplication. The identity elements are denoted by 0 and 1, the additive inverse element by $-a$ and if the multiplicative inverse element exists, it is denoted by a^{-1}. Sometimes rings that are defined like we did, are called unitary rings or rings with 1, but we just call them commutative rings. Moreover, we sometimes want to leave out the word *commutative* for the sake of brevity. So if we talk about a ring, we mean a commutative ring. Non-commutative rings are not considered in this thesis. Sometimes we also just want to write down R instead of $(R, +, \cdot)$. Then the addition and multiplication symbol is assumed to be $+$ and $\cdot$. Sometimes we denote the multiplication also by no symbol. Which symbol we use will be clear from the context.

At first we need some important basic properties that hold in commutative rings.

Lemma 2.2. *Let R be a commutative ring, then for all $r \in R$ holds*

$$r \cdot 0 = 0 \cdot r = 0.$$

Proof. For all $r \in R$ holds

$$r \cdot 0 = r \cdot (0 + 0) = (r \cdot 0) + (r \cdot 0),$$

which means

$$r \cdot 0 = 0.$$

□

Lemma 2.3. *Let R be a commutative ring, then for all $a, b \in R$ holds*

$$a \cdot (-b) = -(a \cdot b).$$

Proof.

$$(a \cdot b) + (a \cdot (-b)) = a \cdot (b + (-b)) = a \cdot 0 = 0$$

by the previous lemma. □

As in the case of groups, we want to define structure preserving types of functions over rings.

Definition 2.6 (ring-homomorphism)**.** Let $(R, +_R, \cdot_R)$ and $(S, +_S, \cdot_S)$ be commutative rings. Then we call a function $\phi : R \to S$ a *(ring-)homomorphism* if ϕ is a group homomorphism over the additive groups $(R, +_R)$ and $(S, +_S)$,

$$\phi(x \cdot_R x') = \phi(x) \cdot_S \phi(x')$$

for all $x, x' \in R$ and

$$\phi(1_R) = 1_S,$$

whereby 1_R and 1_S are the multiplicative identity elements of R and S, respectively.

In an analogous way we define *(ring-)endomorphisms*, *-isomorphisms*, *-monomorphisms* and *-epimorphisms*. Again, we define the **kernel** $\operatorname{Ker}\phi$ of such a ring-homomorphism ϕ as $\operatorname{Ker}\phi := \{x \in R \mid \phi(x) = 0\}$. Note that 0 is the additive identity element of S.

We cannot divide in rings because there is no multiplicative inverse in general. That is the reason why there is another algebraic structure, the so-called *fields*. They are a special form of commutative rings which make assumptions concerning the existence of multiplicative inverses.

Definition 2.7 (field)**.** Let K be a commutative ring. K is called a *field* if $(K \backslash \{0\}, \cdot)$ is an abelian group.

Example. *The integers with common addition and multiplication form a commutative ring, but they are not a field. We just take the integer 2 which has no multiplicative inverse.*

The real numbers, even the rational numbers, with common addition and multiplication are a commutative ring and also a field since every element has a multiplicative inverse.

We know that every element taken from a field has a unique multiplicative inverse. But in some cases we do not need such a strong form of divisibility. Hence, there exist many algebraic structures that approximate divisibility and lie between the ring and the field in the inclusion chain presented in the introduction. The most general form are integral domains, which we will consider in the following. This needs the notion *zero divisor*.

Definition 2.8. Let R be a commutative ring. An element $a \in R$ is called *zero divisor* if there exists $b \in R \setminus \{0\}$ such that

$$a \cdot b = 0.$$

We take a look at zero divisors in fields first.

Example. *Let K be a field. Then only 0 is a zero divisor because $0 \cdot a = 0$ for all $a \in K$. Every other element $x \in K \setminus \{0\}$ is not a zero divisor because it is well-known that*

$$x \cdot y = 0 \Rightarrow x = 0 \vee y = 0.$$

This does not hold in arbitrary rings. Indeed there are rings that have non-trivial zero divisors.

Example. *The commutative ring $\mathbb{Z}/4\mathbb{Z}$ has 2 as zero divisor because $2 \cdot 2 = 0$. By the numbers we mean the modulo equivalence classes of them.*

Commutative rings like $\mathbb{Z}/4\mathbb{Z}$ that have non-trivial zero divisors do not allow one to apply the cancellation law. Indeed, $2{\cdot}2 = 0{\cdot}2$ holds in $\mathbb{Z}/4\mathbb{Z}$, but $2 \neq 0$. Therefore, we want to separately define commutative rings that allow the cancellation law. At this point integral domains become important.

Definition 2.9 (integral domain)**.** Let R be a commutative ring. R is called an *integral domain* if only 0 is a zero divisor of R.

We already know that every field is an integral domain. But there are integral domains that are not a field. We just need to consider the following example.

Example. *The integers $\mathbb{Z}$ together with the common addition and multiplication are an integral domain because*

$$a \cdot b = 0 \Rightarrow a = 0 \vee b = 0.$$

But the integers are not a field because 2 has no multiplicative inverse.

It still needs to be shown that the cancellation law actually holds in integral domains.

Lemma 2.4 (cancellation law)**.** *Let R be an integral domain. Then for all $a, b, c \in R$*

$$a \cdot b = a \cdot c, a \neq 0 \Rightarrow b = c.$$

Proof. We use Lemma 2.3 and know that

$$a \cdot (b + (-c)) = (a \cdot b) + (a \cdot (-c)) = (a \cdot b) + (-(a \cdot c)) = 0$$

which implies $b - c = 0$ because a is not zero and hence no zero divisor. Therefore $b = c$. □

2.1.3 Modules

The concept of modules is analogous to vector spaces. We can say informally that modules are just vector spaces defined over commutative rings. The reason why this has an extra name is that many notions known from vector spaces are no longer well-defined over arbitrary rings. Later on we will see an example of this.

Definition 2.10 (modules)**.** Let R be a commutative ring. We call $(M, \oplus, \odot)$ together with $\oplus : M \times M \to M$ and $\odot : R \times M \to M$ a *module* if $(M, \oplus)$ is an abelian group and for all $r, s \in R, x, y \in M$

- $r \odot (x \oplus y) = (r \odot x) \oplus (r \odot y)$,
- $(r + s) \odot x = (r \odot x) \oplus (s \odot x)$,
- $(r \cdot s) \odot x = r \odot (s \odot x)$ and
- $1 \odot x = x$.

$\oplus$ is called vector addition (or just addition) and $\odot$ scalar multiplication. Again, we just want to write down M instead of $(M, \oplus, \odot)$ sometimes.

We denoted the vector addition by $\oplus$ to distinguish it from the addition in the commutative ring. Because the difference should be clear now, we denote the vector addition by $+$. We also denote $\odot$ by $\cdot$ in the following. Most often we leave out the symbol of the scalar multiplication and just use no symbol. Indeed, we write rx instead of $r \cdot x$.

If R is a field in the above definition, then we call a module also a **vector space** over R or a R-vector space. It is well-known that every vector space has a basis, whether a finite one or an infinite one. This is not true about modules in general. Indeed there are modules that do not have a linearly independent system of vectors that are generating the whole module. An example will be presented after the next few definitions.

Definition 2.11 (linear independence)**.** Let M be a R-module and $S \subseteq M$. S is called *linearly independent* if for all $s_1, ..., s_n \in S$ and $\lambda_1, ..., \lambda_n \in R$ with $n \in \mathbb{N}$ the implication

$$\sum_{j=1}^{n} \lambda_j s_j = 0 \Rightarrow \lambda_1 = ... = \lambda_n = 0$$

applies. Otherwise we call S *linearly dependent*.

If S is a finite set, we can easily verify that it is enough to show the implication for all elements of S with $n = \#S$.

Definition 2.12 (generator)**.** Let M be a R-module and $S \subseteq M$. S is called a *generator* of M if for every $m \in M$ there exist $\lambda_1, ..., \lambda_n \in R$ and $s_1, ..., s_n \in S$ with $n \in \mathbb{N}$ such that

$$m = \sum_{j=1}^{n} \lambda_j s_j.$$

By putting these two notions together, we obtain a basis.

Definition 2.13 (basis)**.** Let M be a R-module and $S \subseteq M$. S is called a *basis* of M if S is linearly independent and a generator of M.

In an analogous way we want to define the notions *linearly independent*, *linearly dependent*, *generator* and *basis* for a family $(s_i)_{i \in I}$ of elements taken from a module M. The reason why we are defining the notions over sets and families is that one option is sometimes more comfortable to use than the other option depending on the situation.

Example. *We said that not every module has a basis. Indeed this is true for the $\mathbb{Z}$-module $\mathbb{Q}$. $(\mathbb{Q}, +)$, whereby $+$ is the common addition on the rational numbers, is an abelian group and the scalar multiplication defined as the common multiplication on the rational numbers also satisfies the required module-axioms such that $\mathbb{Q}$ is a $\mathbb{Z}$-module.*

Now assume that there exists a basis. Every subset of $\mathbb{Q}$ with two elements q_1,q_2 is linearly dependent because $q_1 = \frac{a}{b}$ and $q_2 = \frac{c}{d}$ for some $a, b, c, d \in \mathbb{Z}$ and $(c \cdot b) \cdot \frac{a}{b} + (-a \cdot d) \cdot \frac{c}{d} = 0$. But a set consisting of one element cannot generate every rational number because the coefficient is taken from the integers. Hence a basis cannot exist.

In this case it is not possible to define transformation matrices of endomorphisms over those modules that do not have a basis. Transformation matrices will be discussed in Section 2.3.1. To prepare for this we separately define modules that actually have a basis.

Definition 2.14 (free modules)**.** Let M be a R-module. M is called *free* if M has a basis.

We also want to consider linear mappings defined on modules as they are known on vector spaces. The most relevant mappings will be those that map elements taken from a module into the module itself. We will call them *module-endomorphisms.*

Moreover, Section 2.3.1 will state that every *module-homomorphism* on free finite-dimensional modules can be expressed by a transformation matrix. This makes it possible to prove many of the subsequent assertions over matrices, which is sometimes much easier.

Definition 2.15 (module-homomorphism)**.** Let M and N be R-modules, whereby the addition is denoted by $+_M$ and $+_N$, the scalar multiplication by $\cdot_M$ and $\cdot_N$, and $\phi : M \to N$ be a linear function, that is,

(i) $\phi(m +_M m') = \phi(m) +_N \phi(m')$ (additive) and

(ii) $\phi(\lambda \cdot_M m) = \lambda \cdot_N \phi(m)$ (homogeneous)

for all $m, m' \in M$ and $\lambda \in R$. Then we call ϕ *(module-)homomorphism.*

- If ϕ is bijective, we also call ϕ *(module-)isomorphism* and M and N are called *isomorphic.* Then we write down $M \cong N$.

- If $M = N$, we also call ϕ *(module-)endomorphism.* If the function is additionally bijective, we say *(module-)automorphism* to ϕ.
- If ϕ is injective, we also call ϕ *(module-)monomorphism.*
- If ϕ is surjective, we also call ϕ *(module-)epimorphism.*

The **kernel** $\operatorname{Ker}\phi$ of such a module-homomorphism is defined as $\operatorname{Ker}\phi := \{x \in M \mid \phi(x) = 0\}$. Note that 0 denotes the additive identity element of N.

An example concerning module-isomorphisms is contained in the Section 2.3.1 about transformation matrices, where the coordinate mapping is presented. Also note that $\cong$ is an equivalence relation because every module is isomorphic to itself by the identity mapping, every isomorphism has an isomorphism as inverse mapping and because the composition of isomorphisms is an isomorphism itself.

We know from vector spaces that there is a well-defined dimension. This still holds for free modules over commutative rings.

The dimension of a free module We are able to assign a well-defined dimension to every free module. Therefore, we show that every basis of such a free module has the same cardinality. We need to consider ideals, the existence of a maximum ideal by using the axiom of choice and after that we use this ideal to construct a factor ring that is a field [KM17; Pau18]. This helps us to apply the well-definedness of the dimension of vector spaces and generalize this to modules over commutative rings [Baz10].

We begin with the definition of ideals. Usually, left and right ideals are considered separately. But in the case of commutative rings the notions *left ideal* and *right ideal* are the same.

Definition 2.16 (ring ideal)**.** Let R be a commutative ring. Then we call $I \subseteq R$ an *ideal* of R if

- $0 \in I$,
- $\forall a, b \in I : a - b \in I$ and
- $\forall a \in I, r \in R : r \cdot a \in I$.

The first and second criterion are the well-known subgroup properties. That means, if these are satisfied, we know that $(I, +)$ is a subgroup of the additive abelian group $(R, +)$. If we would have defined rings without the requirement to have the multiplicative identity element $1 \in R$, we could state that every ideal is a subring of R. But in our case we cannot because 1 is no member of I in general. Indeed, consider the following example.

Example. $2\mathbb{Z}$ *together with the common addition and multiplication on the integers is an ideal of* $\mathbb{Z}$ *since*

- $0 = 2 \cdot 0 \in 2\mathbb{Z}$,
- $2 \cdot z_1 - 2 \cdot z_2 = 2 \cdot (z_1 - z_2) \in 2\mathbb{Z}$ *for all* $a = 2 \cdot z_1, b = 2 \cdot z_2 \in 2\mathbb{Z}$ *and*

- $r \cdot 2 \cdot z = 2 \cdot r \cdot z \in 2\mathbb{Z}$ *for all* $r \in \mathbb{Z}, a = 2 \cdot z \in 2\mathbb{Z}$.

But $1 \notin 2\mathbb{Z}$*, therefore this ideal is no subring.*

If we interpret R as a R-module, every ideal of it forms a submodule. Indeed we call a subset $S \subseteq M$ of a module $(M, +, \cdot)$ a **submodule** of M if $(S, +, \cdot)$ is a module by itself.

Next we want to state a fact about the existence of a maximal ideal. It is not obvious that this is true. Indeed an important condition that this holds is that the ring has to contain the multiplicative identity element. In our case this is satisfied anyway since we defined commutative rings like that.

From now on we will call an ideal $M \neq R$ of a commutative ring R **maximal** if only M itself and R are ideals containing M.

Lemma 2.5 (W. Krull)**.** *Let R be a commutative ring. Then for every ideal $I \subsetneq R$ there exists a maximal ideal M of R that contains I, that is, $I \subseteq M$.*

Proof. We consider a set X of all ideals of R that contain I and are unequal to R, namely, $X := \{Y \subsetneq R \mid Y \supseteq I, Y \text{ ideal of } R\}$. $(X, \subseteq)$ is a partially ordered set. Now we take an arbitrary chain C of X, that is, a set $C \subseteq X$ such that we have $A \subseteq B$ or $B \subseteq A$ for all $A, B \in C$.

By defining $D := \bigcup_{A \in C} A$, we obtain an ideal with $I \subseteq D$. We verify this by the following three steps.

- 0 is contained in all ideals taken from X, hence 0 certainly is a member of the union of all these ideals.
- Let $a, b \in D$, then we know that there exist $A, B \in C$ such that $a \in A, b \in B$. Because the chain C is totally ordered, we have $A \subseteq B$ or $B \subseteq A$. Without loss of generality we assume the first case. Then $a, b \in B$ and because B is an ideal, we conclude $a - b \in B \subseteq D$.
- For all $r \in R$ and $a \in D$ holds $a \in A$ for some $A \in C$ and therefore $r \cdot a \in A \subseteq D$.

Moreover, we know that every ideal A in X does not contain 1. Otherwise we could conclude that $r \cdot 1 \in A$ for all $r \in R$, hence $A = R$. But this is a contradiction to the definition of X. So this implies that also D does not contain 1, which means that $D \neq R$, which in turn means $D \in X$.

D is an upper bound of C, which implies that $(X, \subseteq)$ is inductively ordered. Hence, Zorn's lemma states the existence of a maximal ideal M of R with $I \subseteq M$. □

We can define a multiplication on the factor group in such a way that it becomes a ring. Moreover, there exists a canonial ring-epimorphism.

Lemma 2.6. *Let A be an ideal of a commutative ring R.*

(i) The factor group $(R/A, +)$ together with the multiplication

$$(a + A) \cdot (b + A) := ab + A$$

forms a commutative ring.

(ii) $\pi : a \mapsto a + A, a \in R$ *is a ring-epimorphism from* R *to* R/A *with* A *as kernel.*

Proof. The multiplication is well-defined because

$$\begin{aligned} & a + A = a' + A, b + A = b' + A \\ & \Rightarrow \exists x, y \in A : a' = a + x, b' = b + y \\ & \Rightarrow \exists x, y \in A : a'b' - ab = ay + xb + xy \in A \\ & \Rightarrow a'b' + A = ab + A \end{aligned}$$

applies for all $a, a', b, b' \in R$. The ring axioms can be easily verified by using the the ring properties of R. This implies (i).

Now we consider assertion (ii). π is obviously surjective. Moreover, it is a homomorphism which follows from

$$\pi(a + b) = (a + b) + A = (a + A) + (b + A) = \pi(a) + \pi(b)$$

and

$$\pi(ab) = ab + A = (a + A) \cdot (b + A) = \pi(a) \cdot \pi(b)$$

for all $a, b \in R$. Moreover we know for all $a \in A$ that $\pi(a) = a + A = A$, which means $a \in \operatorname{Ker} \pi$ since A is the identity element of the factor group R/A. If we have $k \in \operatorname{Ker} \pi$, we know that $k + A = A$ and due to the fact that A is an ideal, we conclude $k \in A$. Hence, the kernel of π is A and (ii) is proved. □

There is a characterization when such a factor ring is a field. The next lemma formalizes this, whereby we take its proof from the lecture notes on Paulin's course *Introduction to Abstract Algebra* at the University of California, Berkeley [Pau18].

Lemma 2.7. *Let* R *be a commutative ring. An ideal* $M \neq R$ *is maximal if and only if* R/M *is a field.*

Proof. We first prove the implication direction "⇒". Indeed suppose M to be maximal. Now let $b \in R$ with $b \notin M$. The set $B := \{br + a \mid r \in R, a \in M\}$ is an ideal of R that contains M. Hence $B = R$ (since $B \neq M$) and therefore there exist $c \in R$ and $d \in M$ such that $1 = bc + d$. This implies $1 + M = (bc + d) + M = bc + M = (b + M)(c + M)$. Hence R/M is a field. Note that the case $b \in M$ is not important since this would imply $b + M = M$, which is the additive identity element.

To prove the inverse direction assume R/M to be a field and B an ideal that contains M with $M \neq B$. Let $b \in B$ with $b \notin M$. $b + M$ is not the additive identity element M and since R/M is a field there exists a multiplicative inverse $c + M$ with $c \in R$ such that $1 + M = (b + M)(c + M) = bc + M$. This implies $1 - bc \in M \subseteq B$. $(B, +)$ is a group, hence $1 = (1 - bc) + bc \in B$. $r = r1 \in B$ for all $r \in R$ since $1 \in B$ and B is an ideal. We obtain $B = R$ and therefore the maximality of M. □

Now we have reached the point at which we are able to prove that every basis of a free module has the same cardinality. We say that two bases of a module have the same cardinality if they are both finite and have the same number of elements or if they are both infinite sets. Note that we do not distinguish between countable and uncountable infinity. The proof of the following theorem is taken from the lecture notes on Badzioch's abstract algebra course at Buffalo University [Baz10].

Theorem 1. *Let R be a commutative ring and M a free R-module. Then each two bases of M have the same cardinality.*

Proof. Krull's Theorem 2.5 states that there exists a maximal ideal I of R and therefore R/I is a field by the previous lemma. We denote by IM the submodule of M consisting of all finite sums $\sum r_i a_i$ with $r_i \in I$ and $a_i \in M$. It can be easily verified that this is actually a module. It is a subgroup since $0 \in IM$ and the difference of two such sums is again such a sum. Furthermore, the set is closed under scalar multiplication with scalars taken from R since I is an ideal. The factor group M/IM is a R/I-module by defining the scalar multiplication as

$$(r + I) \cdot (x + IM) := rx + IM$$

for all $r + I \in R/I$ and $x + IM \in M/IM$. We actually well-defined the scalar multiplication because for all $r, r' \in R$ and $x, x' \in M$ with $r + I = r' + I$ and $x + IM = x' + IM$ holds

$$rx - r'x' = (r - r')x + r'(x - x') \in IM,$$

which implies

$$rx + IM = r'x' + IM.$$

The other module axioms are verified quickly just by applying definitions. Hence M/IM is a vector space over the field R/I.

Now assume that M has an infinite basis B. In this case we have to show that any other basis of M is also infinite. Indeed we want to conclude by contradiction and assume that there exists a finite basis $B' = \{b'_1, ..., b'_n\}$ of M. We can represent each element of B' as a linear combination of elements taken from B. Let us say we need the finite set $\{b_1, ..., b_n\} \subseteq B$ to represent every element of B'. Now we take an element $b \in B \setminus \{b_1, ..., b_n\}$. This is possible since B is infinite. We know that we can represent b as a linear combination with elements from B' hence also as a linear combination with elements from $\{b_1, ..., b_n\}$. But this is a contradiction to the linear independence of B. Hence B' has to be an infinite set.

Now assume that there exists a finite basis $S = \{b_1, ..., b_n\}$ of M. Then

$$\{b_1 + IM, ..., b_n + IM\}$$

is a basis of the vector space M/IM over R/I. Indeed these vectors are linearly independent since

$$\sum_{j=1}^{n} (\lambda_j + I) \cdot (b_j + IM) = 0,$$

implies

$$\sum_{j=1}^{n} (\lambda_j b_j + IM) = \left(\sum_{j=1}^{n} \lambda_j b_j \right) + IM = 0 + IM,$$

which means $\sum_{j=1}^{n} \lambda_j b_j \in IM$. Since every element of M is a unique linear combination of basis vectors, we know that $\lambda_j \in I$ hence $\lambda_j + I = 0 + I$ for all $j \in \{1, ..., n\}$.

This set of vectors is also a generator of M/IM because for an arbitrary $m + IM$ we obtain $m = \sum_{j=1}^{n} \lambda_j b_j$ for some $\lambda_j \in R$. This means

$$m + IM = \left(\sum_{j=1}^{n} \lambda_j b_j \right) + IM = \sum_{j=1}^{n} (\lambda_j + I) \cdot (b_j + IM).$$

We already know that if one basis of M is finite, every other basis of M has to be finite. Indeed let $S' := \{b'_1, ..., b'_m\}$ be a finite basis of M. Then we know that also $\{b'_1 + IM, ..., b'_m + IM\}$ is a basis of M/IM. But M/IM is a vector space, which implies $m = n$. □

This allows us to well-define the dimension of a module.

Definition 2.17 (dimension of a free module)**.** Let R be a commutative ring and M a free R-module with basis B. If B is a finite set, we define the *dimension* of M as

$$\dim_R M := \#B$$

and say that M is a *finite-dimensional module*. If B is an infinite set, we define the *dimension* of M as

$$\dim_R M := \infty$$

and say that M is an *infinite-dimensional module*. If the underlying commutative ring R is clear, we just write $\dim M$ instead of $\dim_R M$ sometimes.

2.2 Field of fractions of an integral domain

We cannot transfer all notions known from vector spaces to modules. For example, the rank of a matrix over a commutative ring is no longer well-defined, even if we define it over integral domains. To sustain many of these notions, we need to naturally extend an integral domain to a field, the so-called *field of fractions*.

Before we define it, we first remember a well-known example, which demonstrates the concept of the field of fractions.

Example. *We take a look at how the rational numbers are constructed out of the integers. We know that*

$$\mathbb{Q} := \{(a, b) \mid a, b \in \mathbb{Z}, b > 0\} / \sim$$

with the equivalence relation

$$(a, b) \sim (c, d) :\Leftrightarrow ad = bc.$$

$\mathbb{Z}$ *is embedded into* $\mathbb{Q}$ *by the injective mapping*

$$\phi : \mathbb{Z} \to \mathbb{Q},\ \phi(z) := [(z, 1)]_\sim,$$

that is, we can identify $z \in \mathbb{Z}$ *with* $[(z, 1)]_\sim \in \mathbb{Q}$ *and can informally write down* $\mathbb{Z} \subseteq \mathbb{Q}$*. The natural operations* $+$ *and* $\cdot$ *on* $\mathbb{Q}$ *are defined for* $(a, b), (c, d) \in \mathbb{Q}$ *as*

$$[(a, b)]_\sim + [(c, d)]_\sim := [(ad + bc, bd)]_\sim$$

and

$$[(a,b)]_\sim \cdot [(c,d)]_\sim := [(ac,bd)]_\sim.$$

Hence, we constructed the field $\mathbb{Q}$ *out of the integral domain* $\mathbb{Z}$*.*

Now we are going to extend this construction to arbitrary integral domains. Bosch [Bos13] presents a straightforward approach to do this.

Definition 2.18 (field of fractions)**.** Let R be an integral domain. Then we define the *field of fractions* $\mathrm{Frac}(R)$ of R by

$$\mathrm{Frac}(R) := S \,/\, \sim,$$

with $S := \{(a,b) \mid a \in R, b \in R \setminus \{0\}\}$ and $\sim \subseteq S \times S$ with

$$(a,b) \sim (c,d) :\Leftrightarrow ad = bc.$$

Moreover, we define the addition $\oplus$ and multiplication $\odot$ on $\mathrm{Frac}(R)$ by

$$[(a,b)]_\sim \oplus [(c,d)]_\sim := [(ad+bc,bd)]_\sim$$

and

$$[(a,b)]_\sim \odot [(c,d)]_\sim := [(ac,bd)]_\sim.$$

In the following we will denote the equivalence class $[(a,b)]_\sim$ by the fraction symbol $\frac{a}{b}$. Then the definition of the addition and multiplication over the field of fractions look like

$$\frac{a}{b} \oplus \frac{c}{d} = \frac{ad+bc}{bd} \quad \text{and} \quad \frac{a}{b} \odot \frac{c}{d} = \frac{ac}{bd}.$$

Because $\oplus$ and $\odot$ act like the common addition and multiplication of fractions, we will use the symbols $+$ and $\cdot$ from now on.

It remains to verify that this definition actually makes sense. It is not clear that $\sim$ is an equivalence relation, $\oplus$ and $\odot$ are well-defined and $\mathrm{Frac}(R)$ is a field that embeds the integral domain R. Moreover, we need to clarify the neutral element and the inverse of an arbitrary member of the field of fractions.

Lemma 2.8. $\sim$ *is an equivalence relation on* R*.*

Proof. We conclude by verifying symmetry, reflexivity and transitivity. Indeed, let $(a,b),(c,d),(e,f) \in S$.

- From $(a,b) \sim (c,d)$ follows by definition $ad = bc$. Therefore $cb = da$ and again by definition $(c,d) \sim (a,b)$.
- From $ab = ba$ follows immediately $(a,b) \sim (a,b)$.
- Assume $(a,b) \sim (c,d)$ and $(c,d) \sim (e,f)$. Then $ad = bc$ and $cf = de$. We can multiply the first equation by f from the right side, and the second equation by b from the left side. We obtain $adf = bcf$ and $bcf = bde$. Thus, $adf = bde$, which is equivalent to $afd = bed$. Since the cancellation law holds in integral domains, we get the equation $af = be$, which means $(a,b) \sim (e,f)$.

Therefore, the lemma applies. □

Lemma 2.9. *Let R be an integral domain. Then the field of fractions $Frac(R)$ of R is a field together with addition $+$ and multiplication $\cdot$ defined as above. Moreover, $\mathbb{E} : R \to Frac(R)$ defined as $\mathbb{E}(r) := \frac{r}{1}$ is a ring-monomorphism. That is, we obtain an injective embedding of R in $Frac(R)$ such that $\mathbb{E}(0), \mathbb{E}(1)$ are the identity elements in $Frac(R)$, $\mathbb{E}(-r)$ is the additive inverse, $\mathbb{E}(r^{-1})$ is the multiplicative inverse of $\mathbb{E}(r)$ and $\mathbb{E}$ satisfies $\mathbb{E}(r + r') = \mathbb{E}(r) + \mathbb{E}(r')$ and $\mathbb{E}(rr') = \mathbb{E}(r) \cdot \mathbb{E}(r')$.*

Proof. At first we show that the addition $+$ and the multiplication $\cdot$ are well-defined over the field of fractions. We need to show that the value of each operation does not depend on the representatives of the equivalence classes. Therefore, assume $(a, b), (a', b'), (c, d), (c', d') \in S$ with $\frac{a}{b} = \frac{a'}{b'}$ and $\frac{c}{d} = \frac{c'}{d'}$ $(*)$. We have

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} = \frac{a'd' + b'c'}{b'd'} = \frac{a'}{b'} + \frac{c'}{d'},$$

because $(ad+bc)b'd' = adb'd'+bcb'd' = ab'dd'+cd'bb' \overset{(*)}{=} ba'dd'+dc'bb' = bd(a'd'+b'c')$. We also obtain

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd} = \frac{a'c'}{b'd'} = \frac{a'}{b'} \cdot \frac{c'}{d'}$$

by $(ac)(b'd') = (ab')(cd') \overset{(*)}{=} (ba')(dc') = (bd)(a'c')$.

We proceed by showing that the field of fractions is actually a field. $(\mathrm{Frac}(R), +)$ is an abelian group because

- For all $\frac{a}{b}, \frac{c}{d} \in \mathrm{Frac}(R)$ holds $\frac{ad+bc}{bd} \in \mathrm{Frac}(R)$ since $ad+bc \in R$ and $bd \in R\backslash\{0\}$.
- Arbitrary $\frac{a}{b}, \frac{c}{d}, \frac{e}{f} \in \mathrm{Frac}(R)$ fulfill

  $$\left(\frac{a}{b} + \frac{c}{d}\right) + \frac{e}{f} = \frac{(ad + bc)f + (bd)e}{(bd)f} = \frac{a(df) + b(cf + de)}{b(df)} = \frac{a}{b} + \left(\frac{c}{d} + \frac{e}{f}\right)$$

  and

  $$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} = \frac{cb + da}{db} = \frac{c}{d} + \frac{a}{b},$$

  because R is a commutative ring, which means that addition and multiplication are associative and commutative.
- $\frac{0}{1}$ is the identity element of $\mathrm{Frac}(R)$. Indeed, for all $\frac{a}{b} \in \mathrm{Frac}(R)$

  $$\frac{a}{b} + \frac{0}{1} = \frac{a1 + b0}{b1} = \frac{a}{b}.$$

- Let $\frac{a}{b} \in \mathrm{Frac}(R)$. The additive inverse of this element is $\frac{-a}{b}$ because

  $$\frac{a}{b} + \frac{-a}{b} = \frac{ab + b(-a)}{bb} = \frac{ab + (-a)b}{bb} = \frac{ab + (-ab)}{bb} = \frac{0}{bb} = \frac{0}{1},$$

  whereby the rightmost equality follows from $0 \cdot 1 = 0 = (bb)0$. Hence $-\frac{a}{b} = \frac{-a}{b}$.

Also $(\mathrm{Frac}(R) \setminus \{0\}, \cdot)$ is an abelian group. (Note that 0 is the additive identity element $\frac{0}{1}$ now.) This is verified as follows:

- For all $\frac{a}{b}, \frac{c}{d} \in \mathrm{Frac}(R) \setminus \{0\}$ holds $\frac{ab}{cd} \neq 0$ $(= \frac{0}{1})$ because $a, b \neq 0$ implies $ab \neq 0$ since an integral domain has no zero divisors. Together with $ab, cd \in R \setminus \{0\}$ follows $\frac{ab}{cd} \in \mathrm{Frac}(R) \setminus \{0\}$.
- Arbitrary $\frac{a}{b}, \frac{c}{d}, \frac{e}{f} \in \mathrm{Frac}(R)$ fulfill

$$\left(\frac{a}{b} \cdot \frac{c}{d}\right) \cdot \frac{e}{f} = \frac{(ac)e}{(bd)f} = \frac{a(ce)}{b(df)} = \frac{a}{b} \cdot \left(\frac{c}{d} \cdot \frac{e}{f}\right),$$

  whereby we used the associativity that holds in the integral domain R. Also the commutativity directly transfers from the commutativity in the integral domain. Indeed, we have

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd} = \frac{ca}{db} = \frac{c}{d} \cdot \frac{a}{b}.$$

- The identity element is $\frac{1}{1}$ because for all $\frac{a}{b} \in \mathrm{Frac}(R)$ holds

$$\frac{a}{b} \cdot \frac{1}{1} = \frac{a1}{b1} = \frac{a}{b}.$$

- The inverse of $\frac{a}{b} \in \mathrm{Frac}(R) \setminus \{0\}$ is $\frac{b}{a}$ because

$$\frac{a}{b} \cdot \frac{b}{a} = \frac{ab}{ba} = \frac{ab}{ab} = \frac{1}{1}.$$

Moreover the multiplication is distributive over the addition since

$$\begin{aligned}\frac{a}{b} \cdot \left(\frac{c}{d} + \frac{e}{f}\right) &= \frac{a(cf + de)}{b(df)} = \frac{a(cf) + a(de)}{b(df)} = \frac{b(a(cf) + a(de))}{b(b(df))} \\ &= \frac{(ac)(bf) + (bd)(ae)}{(bd)(bf)} = \left(\frac{a}{b} \cdot \frac{c}{d}\right) + \left(\frac{a}{b} \cdot \frac{e}{f}\right)\end{aligned}$$

for arbitrary $\frac{a}{b}, \frac{c}{d}, \frac{e}{f} \in \mathrm{Frac}(R)$. Therefore $(\mathrm{Frac}(R), +, \cdot)$ is a field. It still needs to be shown that $\mathbb{E}$ is a ring-monomorphism. Indeed $\mathbb{E}$ is additive and multiplicative, which follows from

$$\mathbb{E}(r + r') = \frac{r + r'}{1} = \frac{r1 + 1r'}{1} = \frac{r}{1} + \frac{r'}{1} = \mathbb{E}(r) + \mathbb{E}(r')$$

and

$$\mathbb{E}(rr') = \frac{rr'}{1} = \frac{r}{1} \cdot \frac{r'}{1} = \mathbb{E}(r) \cdot \mathbb{E}(r')$$

for all $r, r' \in R$. $\mathbb{E}$ maps the multiplicative identity element 1 of R to the multiplicative identity element $\frac{1}{1}$ of $\mathrm{Frac}(R)$. Moreover from $\frac{r}{1} = \frac{r'}{1}$ follows $r = r'$ by the definition of the equivalence relation $\sim$ and therefore $\mathbb{E}$ is injective. The other properties stated in the lemma are trivial since they hold for arbitrary ring-homomorphisms. But we prove them anyway to get a better understanding: The

identity elements 0 and 1 in R are exactly mapped to the identity elements in $\mathrm{Frac}(R)$. Moreover, if r has the additive inverse $-r$, then $\frac{r}{1} + \frac{-r}{1} = \frac{0}{1}$. Also if $r \in R$ has the multiplicative inverse r^{-1}, we know that $rr^{-1} = 1$, hence $\frac{r}{1} \cdot \frac{r^{-1}}{1} = \frac{1}{1}$. Because $\mathrm{Frac}(R)$ is a field, we know that the inverse elements are unique. Thus, $\frac{-r}{1} = -\frac{r}{1}$ and $\frac{r^{-1}}{1} = \frac{1}{r} = \left(\frac{r}{1}\right)^{-1}$. □

In the following, we will call $\mathbb{E}$ the **embedding** of R in $\mathrm{Frac}(R)$.

2.3 Basics of matrices and homomorphisms

This section is made to give an introduction to basic concepts concerning matrices. For instance, we will get to know what transformation matrices and minimal polynomials are. Meanwhile, we also prove the famous Cayley-Hamilton theorem. These notions and statements are fundamental and necessary for the subsequent content.

We denote by $\mathbb{I}_n$ the $n \times n$ identity matrix having the entries on the main diagonal equal to 1 and the other entries equal to 0. By $\mathbb{O}_{m\times n}$ we mean the $m \times n$ matrix having all entries zero.

Let R be a commutative ring and $A \in R^{m\times n}$. Then we define the **kernel** $\mathrm{Ker}\, A$ of A by $\mathrm{Ker}\, A := \{x \in R^n \mid Ax = 0\}$. The **image** $\mathrm{Im}\, A$ of A is defined by $\mathrm{Im}\, A := \{y \in R^m \mid \exists x \in R^n : Ax = y\}$. If R is a field, the **rank** of a matrix is well-defined and we denote it by $\mathrm{rank}\, A$. Also remember the well-known inequality about the rank: *Let K be a field, $A \in K^{m\times n}$ and $B \in K^{n\times k}$. Then* $\mathrm{rank}(AB) \leq \min\{\mathrm{rank}\, A, \mathrm{rank}\, B\}$.

Now let K be a field and $A = (a_{i,j})_{1\leq i,j\leq n} \in K^{n\times n}$. We define the **characteristic polynomial** P_A of A by $P_A(t) := \det(A - t\mathbb{I}_n) \in K[t]$. Moreover we define

$$A_{ij} := \begin{pmatrix} a_{1,1} & \cdots & a_{1,j-1} & 0 & a_{1,j+1} & \cdots & a_{1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{i-1,1} & \cdots & a_{i-1,j-1} & 0 & a_{i-1,j+1} & \cdots & a_{i-1,n} \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ a_{i+1,1} & \cdots & a_{i+1,j-1} & 0 & a_{i+1,j+1} & \cdots & a_{i+1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,j-1} & 0 & a_{n,j+1} & \cdots & a_{n,n} \end{pmatrix}$$

and

$$a_{ij}^{\#} := \det A_{ji}.$$

Then the **complementary matrix** $A^{\#}$ of A is defined by

$$A^{\#} := (a_{ij}^{\#})_{1\leq i,j\leq n} \in K^{n\times n}.$$

We will need the following result [Fis13] in the proof of the Cayley-Hamilton theorem: *Let $A \in K^{n\times n}$ and $A^{\#}$ the complementary matrix of A. Then*

$$A^{\#}A = AA^{\#} = (\det A)\mathbb{I}_n.$$

Proof. Let $i, j \in \{1, ..., n\}$. Then we have

$$\begin{aligned}
\sum_{k=1}^{n} a_{ik}^{\#} a_{kj} &= \sum_{k=1}^{n} a_{kj} \det A_{ki} \\
&= \sum_{k=1}^{n} a_{kj} \det (a_1, ..., a_{i-1}, e_k, a_{i+1}, ..., a_n) \\
&= \det (a_1, ..., a_{i-1}, \sum_{k=1}^{n} a_{kj} e_k, a_{i+1}, ..., a_n) \\
&= \det (a_1, ..., a_{i-1}, a_j, a_{i+1}, ..., a_n) \\
&= \delta_{ij} \det A,
\end{aligned}$$

whereby δ_{ij} is the Kronecker delta (equals 1 if $i = j$ and equals 0 if $i \neq j$). Hence

$$A^{\#} A = \left(\sum_{k=1}^{n} a_{ik}^{\#} a_{kj} \right)_{1 \leq i,j \leq n} = (\delta_{ij} \det A)_{1 \leq i,j \leq n} = (\det A) \mathbb{I}_n.$$

$AA^{\#} = (\det A)\mathbb{I}_n$ is proved in an analogous way. □

We need one last definition before we consider transformation matrices.

Definition 2.19 (similar matrices)**.** Let R be a commutative ring and $A, B \in R^{n \times n}$. A and B are called *similar* if there exists an invertible matrix $\Gamma \in R^{n \times n}$ such that $A = \Gamma^{-1} B \Gamma$.

At the first glance, it may seem unclear, why the matrices should be "similar" if such an invertible matrix exists. But it will turn out that we can consider a matrix to be a linear mapping over modules and vice versa. Then similarity of matrices means that the matrices belong to the same linear mapping with respect to a different basis.

2.3.1 Transformation matrices

As we have already said, we want to translate endomorphisms over modules into matrices such that we can determine function values by matrix multiplication. We want to do this for arbitrary free finite-dimensional modules over a commutative ring. The problem is that the underlying free R-module of the endomorphism can have an arbitrary structure. Therefore, we need an isomorphic projection to the module $R^{\dim M}$. Then we are able to define a so-called *transformation matrix* that acts on this isomorphic module.

To find such an isomorphism to $R^{\dim M}$ for every module M we need the coordinate mapping. We proceed like Lang [Lan04], whereby we also use some proofs from [Fis13] and generalize them to commutative rings. In the following, R denotes a commutative ring.

Definition 2.20. Let M be a free finite-dimensional R-module with $n := \dim M$ and $B = (b_1, ..., b_n)$ be a basis of M. Then we define the *coordinate mapping* $\Omega_B : M \to R^n$ of M with basis B for all $x \in M$ by

$$\Omega_B(x) = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} :\Leftrightarrow x = \sum_{j=1}^{n} \alpha_j b_j.$$

That is, Ω_B maps every element x to its coordinates with respect to the basis B. We can imagine the basis vectors to be coordinate axes and now every element x is assigned a position in that coordinate system. The module has the same structure as this coordinate system. Indeed, the following lemma formalizes this.

Lemma 2.10. *The coordinate mapping Ω_B is an isomorphism. In particular, M and $R^{\dim M}$ are isomorphic modules.*

Proof. We have to show that Ω_B is a homomorphism $(*)$ and bijective $(\#)$. Let $x, y \in M$ and $\lambda, \alpha_1, ..., \alpha_n, \beta_1, ..., \beta_n \in R$ with $x = \sum_{j=1}^{n} \alpha_j b_j$ and $y = \sum_{j=1}^{n} \beta_j b_j$. Then

$$x + \lambda y = \sum_{j=1}^{n} \alpha_j b_j + \lambda \sum_{j=1}^{n} \beta_j b_j = \sum_{j=1}^{n} (\alpha_j + \lambda \beta_j) b_j,$$

hence

$$\Omega_B(x + \lambda y) = \begin{pmatrix} \alpha_1 + \lambda\beta_1 \\ \vdots \\ \alpha_n + \lambda\beta_n \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} + \lambda \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} = \Omega_B(x) + \lambda \Omega_B(y),$$

which implies $(*)$. Now we show that the coordinate mapping is injective by verifying $\operatorname{Ker} \Omega_B = \{0\}$. Indeed, $\Omega_B(0) = 0 \Rightarrow 0 \in \operatorname{Ker} \Omega_B$ and

$$v \in \operatorname{Ker} \Omega_B \Rightarrow \Omega_B(v) = 0 \Rightarrow v = \sum_{j=1}^{n} 0 \cdot b_j = 0.$$

The surjectivity is obtained as follows: For arbitrary $y = (y_1, ..., y_n)^T \in R^n$, we know that $x := \sum_{j=1}^{n} y_j b_j \in M$ and $\Omega_B(x) = y$. Hence $(\#)$ follows. So Ω_B is an isomorphism on M and $R^{\dim M}$ and these two modules are isomorphic. □

Note that we can use the inverse mapping Ω_B^{-1} from now on.

Definition 2.21 (transformation matrix)**.** Let M, N be free finite-dimensional R-modules of dimension m and n. Further, let $B = (b_1, ..., b_m)$ be a basis of M, $C = (c_1, ..., c_n)$ be a basis of N and $\phi : M \to N$ be a homomorphism. The *transformation matrix* $M_{B,C}(\phi)$ of ϕ is defined by

$$M_{B,C}(\phi) := (\Omega_C(\phi(b_1)), ..., \Omega_C(\phi(b_m))).$$

We can use this transformation matrix in the following way.

Lemma 2.11. *Let M, N be free finite-dimensional R-modules of dimension m and n. Further, let $B = (b_1, ..., b_m)$ be a basis of M, $C = (c_1, ..., c_n)$ be a basis of N and $\phi : M \to N$ be a homomorphism. Then*

$$\phi(x) = \Omega_C^{-1}(M_{B,C}(\phi)\Omega_B(x))$$

for arbitrary $x \in M$.

Proof. Assume $x = \sum_{j=1}^{m} \alpha_j b_j$. Then

$$\begin{aligned}
\phi(x) &= \phi\left(\sum_{j=1}^{m} \alpha_j b_j\right) \\
&= \sum_{j=1}^{m} \alpha_j \phi(b_j) && - \phi \text{ linear} \\
&= \sum_{j=1}^{m} \alpha_j \Omega_C^{-1}(\Omega_C(\phi(b_j))) \\
&= \Omega_C^{-1}\left(\sum_{j=1}^{m} \alpha_j \Omega_C(\phi(b_j))\right) && - \Omega_C^{-1} \text{ linear} \\
&= \Omega_C^{-1}\left(\sum_{j=1}^{m} \alpha_j M_{B,C}(\phi) e_j^{(m)}\right) && - \Omega_C(\phi(b_j)) \ j\text{-th column of } M_{B,C}(\phi) \\
&= \Omega_C^{-1}\left(M_{B,C}(\phi) \sum_{j=1}^{m} \alpha_j e_j^{(m)}\right) \\
&= \Omega_C^{-1}(M_{B,C}(\phi)\Omega_B(x)).
\end{aligned}$$

□

At this point we want to consider two examples.

Example. *Let K be a field and $A = (a_1, ..., a_n) \in K^{n \times n}$. If we define $\phi : K^n \to K^n$ by $\phi(x) := Ax$, we immediately know that ϕ is linear. We determine the transformation matrix with respect to the canonial basis $B = \{e_1^{(n)}, ..., e_n^{(n)}\}$. Indeed we have*

$$\begin{aligned}
M_{B,B}(\phi) &= (\Omega_B(\phi(e_1^{(n)})), ..., \Omega_B(\phi(e_n^{(n)}))) \\
&= (a_1, ..., a_n) \\
&= A.
\end{aligned}$$

The next example is more concrete.

Example. *We define $\phi : \mathbb{R}^2 \to \mathbb{R}$ by $\phi\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) := 3x + 2y$. Indeed this mapping is linear since*

- $\phi\left(\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}x'\\y'\end{pmatrix}\right)=\phi\left(\begin{pmatrix}x+x'\\y+y'\end{pmatrix}\right)=3(x+x')+2(y+y')=(3x+2y)+(3x'+2y')=\phi\left(\begin{pmatrix}x\\y\end{pmatrix}\right)+\phi\left(\begin{pmatrix}x'\\y'\end{pmatrix}\right)$ *and*
- $\phi\left(\lambda\begin{pmatrix}x\\y\end{pmatrix}\right)=3(\lambda x)+2(\lambda y)=\lambda(3x+2y)=\lambda\phi\left(\begin{pmatrix}x\\y\end{pmatrix}\right)$

for all $\begin{pmatrix}x\\y\end{pmatrix},\begin{pmatrix}x'\\y'\end{pmatrix}\in\mathbb{R}^2$ *and* $\lambda\in\mathbb{R}$. *We take* $B:=\left\{\begin{pmatrix}1\\2\end{pmatrix},\begin{pmatrix}3\\0\end{pmatrix}\right\}$ *as basis of* $\mathbb{R}^2$ *and* $\{2\}$ *as the basis of* $\mathbb{R}$. *Then*

$$\begin{aligned}M_{B,\{1\}}&=\left(\Omega_{\{2\}}\left(\phi\left(\begin{pmatrix}1\\2\end{pmatrix}\right)\right),\Omega_{\{2\}}\left(\phi\left(\begin{pmatrix}3\\0\end{pmatrix}\right)\right)\right)\\&=(\Omega_{\{2\}}(7),\Omega_{\{2\}}(9))\\&=(3.5,\,4.5).\end{aligned}$$

So let us check if the transformation matrix works as we want. For example, $\phi\left(\begin{pmatrix}2\\4\end{pmatrix}\right)=6+8=14$. *Now we calculate the value by our transformation matrix: Because* $\begin{pmatrix}2\\4\end{pmatrix}=2\cdot\begin{pmatrix}1\\2\end{pmatrix}+0\cdot\begin{pmatrix}3\\0\end{pmatrix}$, *we have*

$$\begin{aligned}\phi\left(\begin{pmatrix}2\\4\end{pmatrix}\right)&=\Omega_{\{2\}}^{-1}\left(M_{B,\{2\}}\Omega_B\left(\begin{pmatrix}2\\4\end{pmatrix}\right)\right)\\&=\Omega_{\{2\}}^{-1}\left(\begin{pmatrix}3.5&4.5\end{pmatrix}\begin{pmatrix}2\\0\end{pmatrix}\right)\\&=\Omega_{\{2\}}^{-1}(7)\\&=14.\end{aligned}$$

So our transformation matrix seems to be correct.

We are also able to translate a concatenation of homomorphisms into multiplication of the corresponding transformation matrices.

Lemma 2.12. *Let* M,N,P *be free finite-dimensional* R*-modules of dimension* m, n *and* p. *Further, let* $A=(a_1,...,a_m)$ *be a basis of* M, $B=(b_1,...,b_n)$ *be a basis of* N, $C=(c_1,...,c_p)$ *be a basis of* P *and* $\phi:M\to N$, $\psi:N\to P$ *homomorphisms. Then*

$$M_{A,C}(\psi\circ\phi)=M_{B,C}(\psi)\cdot M_{A,B}(\phi).$$

Proof. The assertion follows from

$$\begin{aligned}&M_{B,C}(\psi)\cdot M_{A,B}(\phi)\\&=M_{B,C}(\psi)\cdot(\Omega_B(\phi(a_1)),...,\Omega_B(\phi(a_m)))&&\text{– by Definition 2.21}\\&=(M_{B,C}(\psi)\Omega_B(\phi(a_1)),...,M_{B,C}(\psi)\Omega_B(\phi(a_m)))\\&=(\Omega_C(\psi(\phi(a_1))),...,\Omega_C(\psi(\phi(a_m))))&&\text{– by Lemma 2.11}\\&=(\Omega_C((\psi\circ\phi)(a_1)),...,\Omega_C((\psi\circ\phi)(a_m)))\\&=M_{A,C}(\psi\circ\phi).&&\text{– by Definition 2.21}\end{aligned}$$

□

Now we consider the special case of endomorphisms. Then the transformation matrices are square matrices. The following lemma claims that bijectivity on endomorphisms transfers to invertibility on matrices. That means, we translate automorphisms to invertible matrices.

Lemma 2.13. *Let M be a free finite-dimensional R-module with bases $B = (b_1, ..., b_n)$ and $B' = (b'_1, ..., b'_n)$ and let $\phi : M \to M$ be an automorphism. Then $M_{B,B'}(\phi)$ is an invertible matrix and $M_{B,B'}(\phi)^{-1} = M_{B',B}(\phi^{-1})$.*

Proof. We have

$$\begin{aligned}
& M_{B,B'}(\phi) \cdot M_{B',B}(\phi^{-1}) \\
&= (M_{B,B'}(\phi)\Omega_B(\phi^{-1}(b'_1)), ..., M_{B,B'}(\phi)\Omega_B(\phi^{-1}(b'_n))) && \text{– by Definition 2.21} \\
&= (\Omega_{B'}(\phi(\phi^{-1}(b'_1))), ..., \Omega_{B'}(\phi(\phi^{-1}(b'_n)))) && \text{– by Lemma 2.11} \\
&= (\Omega_{B'}(b'_1), ..., \Omega_{B'}(b'_n)) \\
&= (e_1^{(n)}, ..., e_n^{(n)}) && \text{– by Definition 2.20} \\
&= \mathbb{I}_n
\end{aligned}$$

and in an analogous way $M_{B',B}(\phi^{-1}) \cdot M_{B,B'}(\phi) = \mathbb{I}_n$, hence

$$M_{B,B'}(\phi)^{-1} = M_{B',B}(\phi^{-1}).$$

□

The next lemma states why the definition of similarity actually makes sense.

Lemma 2.14. *Let M be a free finite-dimensional R-module with bases $B = (b_1, ..., b_n)$ and $B' = (b'_1, ..., b'_n)$ and let $\phi : M \to M$ be an endomorphism. Then $M_{B,B}(\phi)$ and $M_{B',B'}(\phi)$ are similar matrices.*

Proof. We have to prove that there exists an invertible $\Gamma \in R^{n \times n}$ such that $M_{B,B}(\phi) = \Gamma^{-1} M_{B',B'}(\phi)\Gamma$. Define

$$\Gamma := (\Omega_{B'}(b_1), ..., \Omega_{B'}(b_n)),$$

that is, Γ is the transformation matrix $M_{B,B'}(\mathrm{Id}_M)$. Because Id_M is bijective, hence an automorphism, we can apply Lemma 2.13, which states that Γ is invertible. Thus, Γ^{-1} exists. It remains to show that Γ satisfies the desired equation. Indeed,

$$\begin{aligned}
\Gamma^{-1} M_{B',B'}(\phi)\Gamma &= M_{B,B'}(\mathrm{Id}_M)^{-1} M_{B',B'}(\phi) M_{B,B'}(\mathrm{Id}_M) \\
&= M_{B',B}(\mathrm{Id}_M^{-1}) M_{B',B'}(\phi) M_{B,B'}(\mathrm{Id}_M) && \text{– by Lemma 2.13} \\
&= M_{B',B}(\mathrm{Id}_M) M_{B',B'}(\phi) M_{B,B'}(\mathrm{Id}_M) \\
&= M_{B,B}(\mathrm{Id}_M \circ \phi \circ \mathrm{Id}_M) && \text{– by Lemma 2.12} \\
&= M_{B,B}(\phi).
\end{aligned}$$

□

2.3.2 Minimal polynomials

The minimal polynomial only exists for matrices with entries taken from a field. If the entries are taken from a commutative ring, and even from an integral domain, we can no longer state the existence and uniqueness of the minimal polynomial. Indeed, the reason for this is that the set of polynomials with coefficients taken from an integral domain is not a principal ideal domain in general. Hence, the ideal that we will define in Definition 2.22 maybe has no generator. But for us it is enough to consider the following for matrices over fields. Indeed we take the approach of Fischer's linear algebra book [Fis13]. Let K be a field throughout the whole section.

We denote by $K[t]$ the set of all polynomials with coefficients in K. Now we need polynomials that act on matrices. This is done by taking a polynomial $P \in K[t]$ of degree m with $P(t) = \alpha_m t^m + ...\alpha_0 t^0$ for $\alpha_0, ..., \alpha_m \in K$. Then we define $\tilde{P} : K^{n\times n} \to K^{n\times n}$ with $\tilde{P}(M) := \alpha_m M^m + ...\alpha_0 M^0$, whereby $M^0 := \mathbb{I}_n$. Because $\tilde{P}$ does the same with the only difference that it acts on matrices, we identify $\tilde{P}$ with P and use the same symbol P for both functions.

As we already said, the minimal polynomial is the generator of an ideal, in fact of the following set.

Definition 2.22. Let $M \in K^{n\times n}$. Then

$$I_M := \{P \in K[t] \mid P(M) = 0\}$$

defines the set of all polynomials that map M to zero.

It is clear that this set is actually an ideal. Indeed, the constant zero mapping is a member of this ideal and if we have $P, P' \in I_M$, then $(P - P')(M) = 0$, hence $P - P' \in M$. So I_M is an additive subgroup of the group of all polynomials with coefficients in K. Also, if we take any $Q \in K[t]$, then we know that $(Q \cdot P)(M) = Q(M) \cdot P(M) = Q(M) \cdot 0 = 0$.

That the generator of this ideal actually exists is verified by the subsequent lemma.

Lemma 2.15. *Let $M \in K^{n\times n}$. There exists a unique $P \in I_M$ such that*

(i) P is monic, that is, $P(\lambda) = \lambda^n + ...$, and

(ii) for every $Q \in I_M$ exists $R \in K[t]$ with $Q = R \cdot P$.

Before we are able to prove this, we need two preliminary statements: a fundamental statement about the division of polynomials and the well-known Cayley-Hamilton theorem. We denote the degree of a polynomial f by $\deg(f)$.

Proposition 2.1. *Let $f, g \in K[t]$ with $g \neq 0$. Then there exist unique $q, r \in K[t]$ such that*

$$f = q \cdot g + r,$$

whereby $\deg(r) < \deg(g)$.

We do not want to prove this in detail. The assertion should be intuitively clear. A detailed proof can be found in Fischer's linear algebra book [Fis13].

Also the proof of the following proposition is taken from this book.

Proposition 2.2 (Cayley-Hamilton theorem)**.** *Let $A \in K^{n\times n}$ and $P_A \in K[t]$ be the characteristic polynomial of A. Then*

$$P_A(A) = \mathbb{O}_{n\times n}.$$

Proof. At first we define

$$B(t) := (A - t\mathbb{I}_n)^T \in K[t]^{n\times n}$$

which is a matrix with entries that are polynomials with coefficients in K. More precise, the entries on the main diagonal are polynomials, all other entries are elements of K. It is the case that

$$\det B(t) = P_A(t) \in K[t]$$

by the definition of the characteristic polynomial. Now we apply the polynomials that are the entries of $B(t)$ on the matrix A. That means

$$B(A) = \begin{pmatrix} a_{11}\mathbb{I}_n - A & a_{21}\mathbb{I}_n & \cdots & a_{n1}\mathbb{I}_n \\ a_{12}\mathbb{I}_n & a_{22}\mathbb{I}_n - A & \cdots & a_{n2}\mathbb{I}_n \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n}\mathbb{I}_n & a_{2n}\mathbb{I}_n & \cdots & a_{nn}\mathbb{I}_n - A \end{pmatrix} \in K[A]^{n\times n}.$$

We are able to multiply $B(A)$ by $n\times 1$ matrices with entries that are matrices having n rows. Therefore we can multiply $B(A)$ by $(e_j)^T_{1\le j\le n}$. Indeed we obtain

$$B(A)\begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} a_{11}e_1 - Ae_1 + a_{21}e_2 + ... + a_{n1}e_n \\ \vdots \\ a_{1n}e_1 + a_{2n}e_2 + ... + a_{nn}e_n - Ae_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

We take the complementary matrix $B^{\#}(t)$ of $B(t)$ and know that

$$B^{\#}(t)B(t) = \det B(t)\mathbb{I}_n = P_A(t)\mathbb{I}_n.$$

Therefore

$$\begin{pmatrix} P_A(A) & & 0 \\ & \ddots & \\ 0 & & P_A(A) \end{pmatrix}\begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} P_A(A)e_1 \\ \vdots \\ P_A(A)e_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

So $P_A(A)e_j = 0$ for all $j \in \{1, ..., n\}$, which means that all columns of $P_A(A)$ are zero columns. Hence $P_A(A) = \mathbb{O}_{n\times n}$. □

Proof of Lemma 2.15. Let $d := \min\{r \in \mathbb{N} \mid \exists P \in I_M : P \neq 0, \deg(P) = r\}$. This minimum exists because the set is bounded below and is not empty since the Cayley-Hamilton theorem holds. Now we choose a monic $P \in I_M$ with $\deg(P) = d$. It exists because we can choose $\tilde{P} \in I_M$ with $\deg(\tilde{P}) = d$ by the definition of d and divide it by its leading coefficient. Moreover there exist $R, T \in K[t]$ with $\deg(T) < \deg(P) = d$ by Proposition 2.1 such that

$$Q = R \cdot P + T$$

for arbitrary $Q \in I_M$. If $T = 0$, we are done. Otherwise

$$T(M) = Q(M) - R(M) \cdot P(M) = 0 - R(M) \cdot 0 = 0$$

implies $T \in I_M$, which is a contradiction to the minimality of d. Thus, it is always $T = 0$.

Only the uniqueness of P remains to show. So assume another $P' \in I_M$ with $P' \neq P$ satisfying (i) and (ii) of Lemma 2.15. We know that $\deg(P') \geq d = \deg(P)$. If $\deg(P) < \deg(P')$ we get a contradiction to (ii) because there exist unique $R, T \in K[t]$ with $\deg(T) < \deg(P')$, such that $P = R \cdot P' + T$, and $P = 0 \cdot P' + P$. Hence, there cannot exist $R \in K[t]$ such that $P = R \cdot P' + 0$. In the other case $\deg(P') = \deg(P)$, we know that the leading terms of P and P' are equal because they are monic polynomials. Hence $1 \leq \deg(P - P') \leq d - 1$ and $(P - P')(M) = 0$, which is a contradiction to the minimality of d. □

Definition 2.23 (minimal polynomial)**.** The P in Lemma 2.15 is called the *minimal polynomial* of M.

It is easy to see that the minimal polynomial of a matrix is the polynomial with minimal degree such that it maps the matrix to zero. Indeed, this is a result of Lemma 2.15's proof, where d is chosen as the minimum. Moreover, the minimal polynomial is unique.

There is a useful result based on the form of the minimal polynomial depending on the invertibility of the underlying matrix [Sti18].

Lemma 2.16. *Let $M \in K^{n \times n}$ and $P(\lambda) = \lambda^m + a_{m-1}\lambda^{m-1} + ... + a_0\lambda^0$ the minimal polynomial of M. Then*

$$M \text{ invertible} \Leftrightarrow a_0 \neq 0.$$

Proof. We prove both directions.

"⇒" Let M be invertible. Then the assumption

$$M^m + ... + a_1 M = \mathbb{O}_{n \times n}$$

leads to

$$M^{-1}(M^m + ... + a_1 M) = \mathbb{O}_{n \times n} = M^{m-1} + ... + a_1 \mathbb{I}_n,$$

whereby the degree of the rightmost polynomial is $m - 1$. This is a contradiction to the minimal degree of P. Therefore we know that

$$\tilde{M} := M^m + ... + a_1 M \neq \mathbb{O}_{n \times n}.$$

Because

$$P(M) = \tilde{M} + a_0 \mathbb{I}_n = \mathbb{O}_{n \times n},$$

a_0 cannot be zero.

"⇐" Now assume $a_0 \neq 0$. By defining

$$M' := -\frac{1}{a_0} M^{m-1} - \frac{a_{m-1}}{a_0} M^{m-2} - ... - \frac{a_1}{a_0} \mathbb{I}_n,$$

we obtain

$$\begin{aligned} MM' &= M(-\frac{1}{a_0}M^{m-1} - ... - \frac{a_1}{a_0}\mathbb{I}_n) \\ &= -\frac{1}{a_0}M^m - ... - \frac{a_1}{a_0}M \\ &= (-\frac{1}{a_0})(M^m + ... + a_1 M) \\ &= \mathbb{I}_n, \end{aligned}$$

hence $M^{-1} = M'$.

□

If two matrices are transformation matrices of the same homomorphism, that is, the matrices are similar, they have the same minimal polynomial [Sti18].

Lemma 2.17. *Let $A, B \in K^{n\times n}$ be similar matrices. Then A and B have the same minimal polynomial.*

Proof. Let $P(\lambda) = a_n\lambda^n + ... + a_0\mathbb{I}_n$ be the minimal polynomial of A and $\Gamma \in GL_n(K)$ such that $B = \Gamma A\Gamma^{-1}$. Then

$$\begin{aligned} P(B) &= P(\Gamma A\Gamma^{-1}) \\ &= a_n(\Gamma A\Gamma^{-1})^n + ... + a_0\mathbb{I}_n \\ &= a_n\Gamma A^n\Gamma^{-1} + ... + a_0\mathbb{I}_n \\ &= \Gamma(a_n A^n + ... + a_0\mathbb{I}_n)\Gamma^{-1} \\ &= \Gamma P(A)\Gamma^{-1} \\ &= \mathbb{O}_{n\times n}. \end{aligned}$$

The degree of P is minimal because if there would be P' with $P'(B) = \mathbb{O}_{n\times n}$ having degree less than n, $P'(A)$ would be also the zero matrix, hence this leads to a contradiction. □

Therefore, we would be also able to define the minimal polynomial of an endomorphism on vector spaces. But this is not necessary in our case.

3 Long products of matrices

At the end of this chapter, we will prove that there exists a subproduct in every long-enough product of matrices, which is pseudo-regular. This requires a lot of preliminary work like statements about pseudo-regular matrices, the tensor product and the exterior product. Reutenauer [Reu80] already proved this for matrices over fields. We want to generalize this result to integral domains. At first glance this may seem not trivial because we cannot keep the definition of pseudo-regularity from [Reu80]. The reason for this is that many notions, occurring in this definition, are not well-defined for matrices with entries taken from an integral domain anymore. For example, the minimal polynomial maybe does not exist or is not unique. Moreover the rank of a matrix is not well-defined anymore.

Therefore, we take another approach and first state the result over fields. After that we use this in combination with the field of fractions to generalize the result to integral domains. This is done in the last section of this chapter. But first we begin with considerations over fields. So let K be a field in the following.

3.1 Pseudo-regular matrices

For the moment we define pseudo-regular matrices like Reutenauer [Reu80] did. Before that we prove the equivalence of five statements. Then we are able to define a pseudo-regular matrix by any of these statements.

Proposition 3.1. *Let $A \in K^{n\times n}$. The following statements are equivalent.*

(i) *A belongs to a subgroup contained in the multiplicative monoid of $K^{n\times n}$.*

(ii) *There exist $B, C \in K^{n\times n}$ with $\operatorname{rank}(B) = \operatorname{rank}(BCB)$ such that $A = CB$.*

(iii) *The kernel and the range of A are complementary subspaces of $K^{n\times n}$.*

(iv) *λ^2 does not divide the minimal polynomial $P(\lambda)$ of A.*

(v) *A is the null-matrix, an invertible matrix or similar to a matrix of the form $\begin{pmatrix} A' & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$ with $A' \in GL_{n'}(K)$ and $0 < n' < n$.*

Proof. The following implications prove the thesis.

"$(i) \Rightarrow (ii)$": Because A belongs to a subgroup contained in $K^{n\times n}$, there exists an inverse A^{-1} of A such that $AA^{-1} = \mathbb{I}_n$. By defining $C := \mathbb{I}_n$ and $B := A$, we obtain

$$A = \mathbb{I}_n A = CB.$$

Moreover we can state

$$\operatorname{rank}(B) \overset{(*)}{=} \operatorname{rank}(A\mathbb{I}_n AA^{-1}) \overset{(\#)}{\leq} \operatorname{rank}(A\mathbb{I}_n A) = \operatorname{rank}(BCB) \overset{(\#)}{\leq} \operatorname{rank}(B)$$

Indeed, (*) follows from $A = A\mathbb{I}_n AA^{-1}$ and (#) from the well-known inequality $\operatorname{rank}(XY) \leq \min\{\operatorname{rank} X, \operatorname{rank} Y\}$. Thus,

$$\operatorname{rank}(B) = \operatorname{rank}(BCB).$$

"$(ii) \Rightarrow (iii)$": Let $B, C \in K^{n\times n}$ such that $A = CB$ and $\operatorname{rank}(B) = \operatorname{rank}(BCB)$. We apply the Rank-nullity theorem and know that $\dim \operatorname{Ker} B = \dim \operatorname{Ker}(BCB)$. Together with $\operatorname{Ker} B \subseteq \operatorname{Ker}(BCB)$ follows the equality of the kernels, namely, $\operatorname{Ker} B = \operatorname{Ker}(BCB)$. Furthermore,

$$\operatorname{Ker}(B) \subseteq \operatorname{Ker}(CB) = \operatorname{Ker}(A) \subseteq \operatorname{Ker}(BA) = \operatorname{Ker}(BCB) \subseteq \operatorname{Ker}(B)$$

applies, hence all these kernels are equal. Now we want to show that $\operatorname{Ker}(A) \cap \operatorname{Im}(A) = \{\mathbb{O}_{n\times 1}\}$. So let $x \in \operatorname{Ker}(A) \cap \operatorname{Im}(A)$. Since x is a member of the image of A, there exists $y \in K^n$ such that $Ay = x$. Then we also have $BAy = Bx = \mathbb{O}_{n\times 1}$ since x is a member of $\operatorname{Ker}(A) = \operatorname{Ker}(B)$. Hence we obtain $y \in \operatorname{Ker}(BA) = \operatorname{Ker} A$, which means $x = Ay = \mathbb{O}_{n\times 1}$. Thus, we conclude that A's image and kernel only contain the zero vector hence are complementary subspaces.

"$(iii) \Rightarrow (v)$": If $\operatorname{Ker}(A) = \{\mathbb{O}_{n\times 1}\}$, A is invertible. In case $\operatorname{Im}(A) = \{\mathbb{O}_{n\times 1}\}$, A equals $\mathbb{O}_{n\times n}$. Otherwise we define an endomorphism $\tilde{\Phi} : \operatorname{Im}(A) \to \operatorname{Im}(A)$ with $\tilde{\Phi}(v) := Av$. Further, let $\tilde{B} := (\tilde{b}_1, \tilde{b}_2, \ldots, \tilde{b}_{n'})$ be a basis of $\operatorname{Im}(A)$, whereby $n' := \dim(\operatorname{Im}(A))$ and $0 < n' < n$. A', defined as the transformation matrix $M_{\tilde{B},\tilde{B}}(\tilde{\Phi})$, is an invertible matrix because

$$\operatorname{Ker}(\tilde{\Phi}) = \operatorname{Ker}(A) \cap \operatorname{Im}(A) = \{\mathbb{O}_{n\times 1}\}$$

implies that $\tilde{\Phi}$ is an automorphism and we can apply Lemma 2.13.

Now let Φ be an endomorphism on K^n defined by $\Phi(v) := Av$. Further, let $(b_1, b_2, \ldots, b_m)$ a basis of $\operatorname{Ker}(A)$ with $m := \dim \operatorname{Ker}(A)$. Because $\operatorname{Im}(A)$ and $\operatorname{Ker}(A)$ are complementary subspaces, $m = n - n'$ and we can construct a basis $B := (\tilde{b}_1, \ldots, \tilde{b}_{n'}, b_1, \ldots, b_m)$ for K^n. We obtain the transformation matrix

$$\begin{aligned}
M_{B,B}(\Phi) &= (\Omega_B(\Phi(\tilde{b}_1)), \ldots, \Omega_B(\Phi(\tilde{b}_{n'})), \Omega_B(\Phi(b_1)), \ldots, \Omega_B(\Phi(b_m))) \\
&= \left(\begin{pmatrix} \Omega_{\tilde{B}}(\tilde{\Phi}(\tilde{b}_1)) \\ \mathbb{O}_{m\times 1} \end{pmatrix}, \quad \ldots \ , \quad \begin{pmatrix} \Omega_{\tilde{B}}(\tilde{\Phi}(\tilde{b}_{n'})) \\ \mathbb{O}_{m\times 1} \end{pmatrix}, \quad \mathbb{O}_{n\times m} \right) \\
&= \begin{pmatrix} M_{\tilde{B},\tilde{B}}(\tilde{\Phi}) & \mathbb{O}_{n'\times m} \\ \mathbb{O}_{m\times n'} & \mathbb{O}_{m\times m} \end{pmatrix} \\
&= \begin{pmatrix} A' & \mathbb{O}_{n'\times m} \\ \mathbb{O}_{m\times n'} & \mathbb{O}_{m\times m} \end{pmatrix} =: \hat{A}.
\end{aligned}$$

From the definition of Φ follows that also A is a transformation matrix of Φ (with the canonial basis). Hence, we can apply Lemma 2.14, which states the similarity of A and $\hat{A}$.

"$(v) \Rightarrow (iv)$": (v) allows the following case distinction: If $A = \mathbb{O}_{n\times n}$, the minimal polynomial of A is $P(\lambda) = \lambda$. λ^2 does not divide it. If A is invertible, the minimal polynomial P of A satisfies $P(0) \neq 0$ by Lemma 2.16. Therefore, when we divide

$P(\lambda)$ by λ^2, we do not get a polynomial. In the remaining case, let $A' \in GL_{n'}(K)$ be a matrix such that A is similar to $\hat{A} := \begin{pmatrix} A' & \mathbb{O}_{n' \times m} \\ \mathbb{O}_{m \times n'} & \mathbb{O}_{m \times m} \end{pmatrix}$ for $0 < n' < n$, whereby $m := n - n'$, and Q be the minimal polynomial of A'. Then $P(\lambda) := \lambda Q(\lambda)$ is the minimal polynomial of $\hat{A}$. We can verify this as follows:

- P is a polynomial with degree $q + 1$, whereby q is the degree of Q.
- $P(\hat{A}) = \hat{A}Q(\hat{A}) = \begin{pmatrix} A' & \mathbb{O}_{n' \times m} \\ \mathbb{O}_{m \times n'} & \mathbb{O}_{m \times m} \end{pmatrix} \begin{pmatrix} Q(A') & \mathbb{O}_{n' \times m} \\ \mathbb{O}_{m \times n'} & \alpha_0 \mathbb{I}_m \end{pmatrix} \overset{Q(A')=\mathbb{O}_{n' \times n'}}{=} \mathbb{O}_{n \times n}$ with $Q(0) = \alpha_0 \neq 0$.
- The degree of P is minimal. To verify this, assume the opposite: There exists a monic polynomial S such that the degree of S is smaller than the degree of P and $S(\hat{A}) = \mathbb{O}_{n \times n}$. If the degree of S is also less than the degree of Q, we get a contradiction to the minimality of the degree of Q because $S(\hat{A}) = \mathbb{O}_{n \times n}$ implies $S(A') = \mathbb{O}_{n' \times n'}$. Otherwise the degree of S equals the degree of Q. Therefore S must be the same polynomial as Q because the minimal polynomial is unique by Lemma 2.15. This is a contradiction because from $Q(0) \neq 0$ follows that Q cannot be the minimal polynomial of a singular matrix like $\hat{A}$.

Lemma 2.17 states that similar matrices have the same minimal polynomial. Hence P is also the minimal polynomial of A. With $\alpha_0, \ldots, \alpha_q \in K$ and $Q(\lambda) = \sum_{i=0}^{q} \alpha_i \lambda^i$ we can state

$$P(\lambda) = \sum_{i=0}^{q} \alpha_i \lambda^{i+1}$$

$$\Rightarrow \lambda^{-2} P(\lambda) = \sum_{i=0}^{q} \alpha_i \lambda^{i-1} = \alpha_0 \lambda^{-1} + \sum_{i=1}^{q} \alpha_i \lambda^{i-1}.$$

So λ^2 does not divide the minimal polynomial of A due to $\alpha_0 \neq 0$.

"$(iv) \Rightarrow (v)$": Let P be the minimal polynomial of A satisfying (iv). Let us first consider the trivial cases. If P is the identity, A is the null matrix. In the other case $P(0) \neq 0$, we know by Lemma 2.16 that A is invertible.
Now we look at the non-trivial case: From $P(0) = 0$ follows that there exists a polynomial Q of the form $Q(\lambda) = \sum_{i=0}^{m} \alpha_i \lambda^i$ with $P(\lambda) = \lambda Q(\lambda)$ and $Q(0) \neq 0$ (otherwise λ^2 would divide $P(\lambda)$). We can state

$$K^n = \mathrm{Ker}(A) \oplus \mathrm{Ker}(Q(A)).$$

This is a result of:

- $\mathrm{Ker}(A) \cap \mathrm{Ker}(Q(A)) = \{\mathbb{O}_{n\times 1}\}$ because

$$\begin{aligned}
& v \in \mathrm{Ker}(A) \cap \mathrm{Ker}(Q(A)) \\
& \Rightarrow Av = \mathbb{O}_{n\times 1} \wedge Q(A)v = \mathbb{O}_{n\times 1} \\
& \Rightarrow \left(\sum_{i=0}^{m} \alpha_i A^i\right) v = \alpha_0 \mathbb{I}_n v + \overbrace{\sum_{i=1}^{m} \alpha_i A^i v}^{=\mathbb{O}_{n\times 1}} = \mathbb{O}_{n\times 1} \\
& \Rightarrow \alpha_0 \mathbb{I}_n v = \alpha_0 v = \mathbb{O}_{n\times 1} \\
& \overset{\alpha_0 \neq 0}{\Rightarrow} v = \mathbb{O}_{n\times 1}.
\end{aligned}$$

- The linear independence of the basis vectors of $\mathrm{Ker}(A)$ and $\mathrm{Ker}(Q(A))$ is an implication of this. Thus, we only need to prove that

$$\dim \mathrm{Ker}(Q(A)) = \dim \mathrm{Im}(A)$$

since the Rank-nullity theorem holds. Indeed the inequality $\dim \mathrm{Ker}(Q(A)) \leq \dim \mathrm{Im}(A)$ is trivial because the dimension of two complementary subspaces cannot be greater than n. The inverse inequality $\dim \mathrm{Im}(A) \leq \dim \mathrm{Ker}(Q(A))$ follows from $\mathrm{Im}(A) \subseteq \mathrm{Ker}(Q(A))$, which in turn is a result of

$$\begin{aligned}
& v \in \mathrm{Im}(A) \\
& \Rightarrow \exists w \in K^n : Aw = v \\
& \Rightarrow Q(A)v = Q(A)Aw = P(A)w = \mathbb{O}_{n\times n} w = \mathbb{O}_{n\times 1} \\
& \Rightarrow v \in \mathrm{Ker}(Q(A)).
\end{aligned}$$

Finally, by change of basis (like we did in $(iii) \Rightarrow (v)$), we obtain the similarity of A and a matrix of the form $\begin{pmatrix} A' & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$.

"$(v) \Rightarrow (i)$": If A is the null matrix, we know that

$$A \in \{\mathbb{O}_{n\times n}\},$$

whereby $(\mathbb{O}_{n\times n}, \cdot)$ is a group. $\mathbb{O}_{n\times n}$ is self-inverse and is the neutral element.

In case A is invertible, we can state

$$A \in GL_n(K),$$

which is a group.

In the remaining case let $m := n - n'$ and A is a member of the group

$$\left\{ \begin{pmatrix} A' & \mathbb{O}_{n'\times m} \\ \mathbb{O}_{m\times n'} & \mathbb{O}_{m\times m} \end{pmatrix} \middle| \; A' \in GL_{n'}(K) \right\}.$$

The neutral element of this group is $\begin{pmatrix} \mathbb{I}_{n'} & \mathbb{O}_{n'\times m} \\ \mathbb{O}_{m\times n'} & \mathbb{O}_{m\times m} \end{pmatrix}$. The inverse of an element $\begin{pmatrix} A' & \mathbb{O}_{n'\times m} \\ \mathbb{O}_{m\times n'} & \mathbb{O}_{m\times m} \end{pmatrix}$ is $\begin{pmatrix} (A')^{-1} & \mathbb{O}_{n'\times m} \\ \mathbb{O}_{m\times n'} & \mathbb{O}_{m\times m} \end{pmatrix}$. □

Definition 3.1 (Pseudo-regular matrix)**.** A matrix $A \in K^{n\times n}$ is called *pseudo-regular* if A satisfies one condition of Proposition 3.1.

Condition (v) explains well why the term 'pseudo-regular' actually makes sense and is an approximation of invertibility. Indeed we want to take a look at a matrix that is pseudo-regular.

Example. *We consider the matrix* $A := \begin{pmatrix} 1 & 0 & 5 \\ 0 & 0 & 0 \\ 2 & 0 & 3 \end{pmatrix}$*. We can see that we would obtain a matrix having all entries zero in the last row and column if we would exchange the second and third row and after that the second and third column or vice versa. We know that this can be done by multiplying the matrix by* $\Gamma := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ *from the left and from the right. Luckily* Γ *is self-inverse, therefore*

$$\Gamma A \Gamma^{-1} = \begin{pmatrix} 1 & 5 & 0 \\ 2 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix} =: D,$$

which means that A *is pseudo-regular by condition* (v)*. Also the other conditions are satisfied:* (i) *is clear by the proof of Proposition 3.1.* (ii) *is true by choosing* $C = \Gamma^{-1} = \Gamma$ *and* $B = D\Gamma$ *because* $A = CB$ *and* $\operatorname{rank}(BCB) = \operatorname{rank}(D\Gamma\Gamma D\Gamma) = \operatorname{rank}(DD\Gamma) = 2 = \operatorname{rank}(D\Gamma) = \operatorname{rank} B$*. Condition* (iii) *holds: Let* $x, y, z \in \mathbb{R}$*, then*

$$\begin{pmatrix} 1 & 0 & 5 \\ 0 & 0 & 0 \\ 2 & 0 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Leftrightarrow x + 5z = 0 \ \wedge \ 2x + 3z = 0,$$

which is satisfied if and only if $x = 0 = z$ *and* $y \in \mathbb{R}$*. Hence*

$$\operatorname{Ker} A = \left\{ \begin{pmatrix} 0 \\ y \\ 0 \end{pmatrix} \middle| \ y \in \mathbb{R} \right\}.$$

Moreover

$$\begin{pmatrix} 1 & 0 & 5 \\ 0 & 0 & 0 \\ 2 & 0 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x + 5z \\ 0 \\ 2x + 3z \end{pmatrix},$$

hence

$$\operatorname{Im} A = \left\{ x \cdot \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} + y \cdot \begin{pmatrix} 5 \\ 0 \\ 3 \end{pmatrix} \middle| \ x, y \in \mathbb{R} \right\}.$$

So the only vector that is a member of the image and the kernel of A *is the zero vector. Hence the kernel and the image of* A *are complementary subspaces.* (iv) *is*

also true: We first determine the characteristic polynomial of A, namely,

$$\begin{aligned}\det(\lambda I_3 - A) &= \det\begin{pmatrix}\lambda-1 & 0 & 5\\ 0 & \lambda & 0\\ 2 & 0 & \lambda-3\end{pmatrix}\\ &= \lambda((\lambda-1)(\lambda-3)-10)\\ &= \lambda(\lambda-(2+\sqrt{11}))(\lambda-(2-\sqrt{11})).\end{aligned}$$

We know that the minimal polynomial must be a factor of this polynomial. The factor of smallest degree that maps A to zero is

$$P(M) = (M-I_3)(M-3I_3) - 10I_3 = M^2 - 4M - 7M^0,$$

whereby the polynomial now acts on matrices. P is the minimal polynomial and does obviously not divide M^2. Hence A satisfies all conditions in Proposition 3.1.

Another interesting fact is that every symmetric $n \times n$ matrix is pseudo-regular since it has n real eigenvalues and can be diagonalized with these eigenvalues on the diagonal.

3.2 Factors of a matrix-product

We need to specify, what a *factor* of a matrix-product is. For the sake of simplicity, we think of a matrix-product as a family of matrices.

Definition 3.2 (Factor of a matrix-family)**.** Let $n, N \in \mathbb{N}$, $A_1, ..., A_N \in K^{n\times n}$ and $P := (A_k)_{1\le k\le N}$. Then F is called a *factor* of P if there exists $j \in \{1, ..., N\}$ such that $F = (B_k)_{1\le k\le j}$ and for all $1 \le k \le j$ there exists $b_k \in \{1, ..., N\}$ and $b_{j+1} \in \{2, ..., N+1\}$ such that $b_1 < b_2 < ... < b_{j+1}$ and $B_k = A_{b_k}...A_{b_{k+1}-1}$.

We can already state a thesis about the existence of a factor of a matrix-product, satisfying a specific condition about the rank. The proof is based on the idea of Reutenauer [Reu80], who showed the result in a more general way. Reutenauer considered words contained in the freely generated monoid over an alphabet and a mapping r of these words to numbers that satisfies $r(uvw) \le r(v)$. But we prove the result directly over matrices together with the rank that is a function satisfying the required condition $\operatorname{rank}(UVW) \le \operatorname{rank}(V)$.

Lemma 3.1. *Let $k_0, k_1, ..., k_n \in \mathbb{N}$. For each family $P = (A_1, A_2, ..., A_m)$ of square-matrices $A_i \in K^{n\times n}$, $1 \le i \le m$ with $m \ge k_0k_1...k_n$, there exists $l \in \{0, ..., n\}$ such that P has a factor $(B_1, ..., B_{k_l})$ satisfying*

$$\forall i, j \in \{1, ..., k_l\} : \operatorname{rank}(B_i...B_j) = l.$$

Proof. At first we define $r_i := \prod_{j=i}^{n} k_j$.

$(*)$ We show that for all $l \in \{0, ..., n\}$ and every matrix-family $P = (A_1, ..., A_m)$ $(A_i \in K^{n\times n}, 1 \le i \le m)$ with $\operatorname{rank}(A_1...A_m) = l$ and $m \ge r_l$, there exist $\hat{l} \in \{0, ..., n\}$ and a factor $(B_1, ..., B_{k_{\hat{l}}})$ of P that satisfies

$$\forall i, j \in \{1, ..., k_{\hat{l}}\} : \operatorname{rank}(B_i...B_j) = \hat{l}. \tag{1}$$

We prove this by induction.

- First we show $(*)$ for $l = n$. So let $P = (A_1, A_2, ..., A_m)$ with $\text{rank}(A_1A_2...A_m) = n$ and $m \geq r_n$. Then for all $i, j \in \{1, ..., m\}$ with $i \leq j$

$$\begin{aligned} & n = \text{rank}(A_1A_2...A_m) \leq \text{rank}(A_i...A_j) \leq n \\ & \Rightarrow \text{rank}(A_i...A_j) = n. \end{aligned}$$

  So we can choose $(A_1, A_2, ..., A_{k_n})$ as the factor satisfying (1).

- Now let $0 \leq l < n$ and suppose that $(*)$ is true for $l+1, l+2, ..., n$. Let $P = (A_1, ..., A_m)$ with $\text{rank}(A_1A_2...A_m) = l$ and $m \geq r_l = k_l r_{l+1}$. Hence P has a factor

$$B := (\underbrace{B_1, ..., B_{r_{l+1}}}_{\hat{B}_1 := B_1...B_{r_{l+1}}}, \underbrace{B_{r_{l+1}+1}, ..., B_{2r_{l+1}}}_{\hat{B}_2 = B_{r_{l+1}+1}...B_{2r_{l+1}}}, ..., ..., \underbrace{B_{(k_l-1)r_{l+1}+1}, ..., B_{k_l r_{l+1}}}_{\hat{B}_{k_l} = B_{(k_l-1)r_{l+1}+1}...B_{k_l r_{l+1}}}).$$

  If there exists $i \in \{1, ..., k_l\}$ such that $\text{rank}(\hat{B}_i) \geq l + 1 > l$, we can conclude by induction. Otherwise, if $\text{rank}(\hat{B}_i) \leq l$ for all $i \in \{1, ..., k_l\}$, we can state for all $i, j \in \{1, ..., k_l\}$ with $i \leq j$

$$\begin{aligned} & l = \text{rank}(A_1A_2...A_m) \leq \text{rank}(\hat{B}_i...\hat{B}_j) \leq \text{rank}(\hat{B}_i) \leq l \\ & \Rightarrow \text{rank}(\hat{B}_i...\hat{B}_j) = l. \end{aligned}$$

  Hence we can choose $(\hat{B}_1, \hat{B}_2, ..., \hat{B}_{k_l})$ as the factor satisfying (1).

Thus, for every matrix-family $P = (A_1, A_2, ..., A_m)$ with $A_i \in K^{n \times n}$, $1 \leq i \leq m$, with $m \geq k_0 k_1 ... k_n$, we can find $l \in \{0, ..., n\}$ that satisfies the required conditions of the assertion.

□

3.3 Tensor product

The main idea of the tensor product is to transform the Cartesian product of vector spaces into one single vector space, the so-called *tensor product* of these vector spaces. This is done in such a way that every multilinear function with the Cartesian product as domain can be transferred into a linear function with the tensor product as domain, which does the same. We will state in Theorem 2 what this exactly means.

The following introduction to this topic is taken from Fischer's linear algebra book [Fis13].

Definition 3.3 (multilinear functions)**.** Let $V_1, V_2, ..., V_k, W$ be vector spaces over K. A function

$$\xi : V_1 \times V_2 \times ... \times V_k \to W$$

is called *multilinear* (or *k-fold linear*) if for all $i \in \{1, ..., k\}$ and fixed $v_j \in V_j$ $(j \in \{1, ..., i-1, i+1, ..., k\})$ the function

$$\xi_i : V_i \to W, \xi_i(v) = \xi(v_1, ..., v_{i-1}, v, v_{i+1}, ..., v_k)$$

is linear.

Example. *$K[t]_d$ denotes the K-vector space of all polynomials of degree less than or equal to d. Suppose the following function*

$$\xi : \underbrace{K[t]_d \times K[t]_d \times ... \times K[t]_d}_{k \text{ times}} \to K[t]_{kd},\ \xi(p_1, p_2, ..., p_k) := p_1 p_2 ... p_k.$$

ξ is k-fold linear. Indeed, let $i \in \{1, ..., k\}$ and choose arbitrary $p_j \in K[t]_d$ for all $j \in \{1, ..., i-1, i+1, ...k\}$, then

$$\begin{aligned}\xi_i(\alpha p + q) &= \xi(p_1, ..., p_{i-1}, \alpha p + q, p_{i+1}, ..., p_k)\\ &= p_1 ... p_{i-1}(\alpha p + q) p_{i+1} ... p_k\\ &= p_1 ... p_{i-1}(\alpha p) p_{i+1} ... p_n + p_1 ... p_{i-1} q p_{i+1} ... p_k\\ &= \alpha(p_1 ... p_{i-1} p p_{i+1} ... p_n) + p_1 ... p_{i-1} q p_{i+1} ... p_k\\ &= \alpha \xi_i(p) + \xi_i(q)\end{aligned}$$

for all $\alpha \in K$ and $p, q \in K[t]_d$.

It is enough to determine the image of all combinations of the basis vectors of $V_1, ..., V_k$ to define a multilinear function. This is stated by the next lemma.

Lemma 3.2. *Let $V_1, ..., V_k$ be K-vector spaces with bases $(v_i^{(j)})_{i \in I_j}$ of V_j for $j \in \{1, ..., k\}$. Further, let also W be a K-vector space, then for an arbitrary family*

$$(w_{i_1, ..., i_k})_{(i_1, ..., i_k) \in I_1 \times ... \times I_k}$$

in W there exists exactly one multilinear function

$$\xi : V_1 \times ... \times V_k \to W \text{ with } \xi(v_{i_1}^{(1)}, ..., v_{i_k}^{(k)}) = w_{i_1, ..., i_k} \tag{2}$$

for all $(i_1, ..., i_k) \in I_1 \times ... \times I_k$.

Remark. We want to prove the lemma also for infinite-dimensional vector spaces. Let $(v_i)_{i \in I}$ be the basis of a vector space V. In the following, the notion

$$v = \sum_j{}' \lambda_j v_j$$

means for the sake of simplicity that there exist $m \in \mathbb{N}$, $i_1, ..., i_m \in I$ and $\lambda_1, ..., \lambda_m \in K$ such that

$$v = \sum_{j=1}^{m} \lambda_j v_{i_j}.$$

That means v is a linear combination of finitely many basis vectors. Note that if I is finite, that is, V is finite-dimensional, $\sum'$ and $\sum$ are the same.

Proof. We conclude in two steps. First we show the uniqueness of this multilinear function. After that we prove that such a function actually exists.

- Let ξ be a multilinear function that satisfies (2).

$$\begin{aligned}\xi(v_1,...,v_k) &= \xi\left(\sum_{i_1}{}' \lambda_{1,i_1} v_{i_1}^{(1)},...,\sum_{i_k}{}' \lambda_{k,i_k} v_{i_k}^{(k)}\right)\\ &= \sum_{i_1}{}' ... \sum_{i_k}{}' \lambda_{1,i_1}...\lambda_{k,i_k}\xi(v_{i_1}^{(1)},...,v_{i_k}^{(k)})\\ &= \sum_{i_1}{}' ... \sum_{i_k}{}' \lambda_{1,i_1}...\lambda_{k,i_k} w_{i_1,...,i_k}\end{aligned}$$

  applies for all $(v_1,...,v_k) \in V_1 \times ... \times V_k$ with $v_j = \sum_{i_j}{}' \lambda_{j,i_j} v_{i_j}^{(j)}$, $j \in \{1,...,k\}$. Hence there cannot exist any other multilinear function than ξ.

- Now we have to show that ξ defined as

$$\xi(v_1,...,v_k) := \sum_{i_1}{}' ... \sum_{i_k}{}' \lambda_{1,i_1}...\lambda_{k,i_k} w_{i_1,...,i_k}$$

  actually is a multilinear function. We do this by choosing arbitrary and fixed $v_1,...,v_{j-1},v_{j+1},...,v_k$ for any $j \in \{1,...,k\}$.

$$\xi_j : V_j \to W,\ \xi_j(v) := \xi(v_1,...,v_{j-1},v,v_{j+1},...,v_k)$$

  is a linear function. Indeed, for arbitrary $v = \sum_{i_j}{}' \lambda_{i_j}^{(v)} v_{i_j}^{(j)}$ and $w = \sum_{i_j}{}' \lambda_{i_j}^{(w)} v_{i_j}^{(j)}$ taken from V_j, there is

$$\begin{aligned}\xi_j(\alpha v + w) &= \xi_j\left(\alpha \sum_{i_j}{}' \lambda_{i_j}^{(v)} v_{i_j}^{(j)} + \sum_{i_j}{}' \lambda_{i_j}^{(w)} v_{i_j}^{(j)}\right)\\ &= \xi_j\left(\sum_{i_j}{}' (\alpha\lambda_{i_j}^{(v)} + \lambda_{i_j}^{(w)}) v_{i_j}^{(j)}\right)\\ &= \sum_{i_1}{}' ... \sum_{i_k}{}' \lambda_{1,i_1}...\lambda_{j-1,i_{j-1}}(\alpha\lambda_{i_j}^{(v)} + \lambda_{i_j}^{(w)})\lambda_{j+1,i_{j+1}}...\lambda_{k,i_k} w_{i_1,...,i_k}\\ &= \alpha \sum_{i_1}{}' ... \sum_{i_k}{}' \lambda_{1,i_1}...\lambda_{j-1,i_{j-1}}\lambda_{i_j}^{(v)}\lambda_{j+1,i_{j+1}}...\lambda_{k,i_k} w_{i_1,...,i_k} +\\ &\quad \sum_{i_1}{}' ... \sum_{i_k}{}' \lambda_{1,i_1}...\lambda_{j-1,i_{j-1}}\lambda_{i_j}^{(w)}\lambda_{j+1,i_{j+1}}...\lambda_{k,i_k} w_{i_1,...,i_k}\\ &= \alpha\xi(v_1,...,v_{j-1},v,v_{j+1},...,v_k) + \xi(v_1,...,v_{j-1},w,v_{j+1},...,v_k)\\ &= \alpha\xi_j(v) + \xi_j(w).\end{aligned}$$

□

Now we have the tools to prove the existence of the tensor product.

Theorem 2. *Let $V_1, ..., V_k$ be K-vector spaces. Then there exist a vector space $V_1 \otimes ... \otimes V_k$ and a multilinear function*

$$\eta : V_1 \times ... \times V_k \to V_1 \otimes ... \otimes V_k,\ \eta(v_1, ..., v_k) = v_1 \otimes ... \otimes v_k,$$

which have the following universal property: For all K-vector spaces W together with a multilinear function

$$\xi : V_1 \times ... \times V_k \to W,$$

there exists exactly one linear function

$$\xi_\otimes : V_1 \otimes ... \otimes V_k \to W$$

such that $\xi = \xi_\otimes \circ \eta$. If for all $i \in \{1, ..., k\}$ V_i is finite-dimensional with basis

$$(v_1^{(i)}, ..., v_{r_i}^{(i)}),$$

then

$$(v_{i_1}^{(1)} \otimes ... \otimes v_{i_k}^{(k)})_{\forall j \in \{1,...,k\} : 1 \le i_j \le r_j}$$

is a basis of $V_1 \otimes ... \otimes V_k$. In particular,

$$\dim(V_1 \otimes ... \otimes V_k) = \dim V_1 \cdot ... \cdot \dim V_k.$$

Proof. The idea of this proof comes from [SG16]. Let $(v_i^{(j)})_{i \in I_j}$ be the basis of V_j for $j \in \{1, ..., k\}$. We define $V_1 \otimes ... \otimes V_k$ by

$$\{\tau \in K^{\{I_1 \times ... \times I_k\}} : \tau(i_1, ..., i_k) \neq 0 \text{ for only finitely many } (i_1, ..., i_k) \in I_1 \times ... \times I_k\}.$$

It is easy to see that this is a vector space. But the definition itself may seem very abstract. To get a notion of $V_1 \otimes ... \otimes V_k$, we can think of the set containing all linear combinations of formal expressions that have the form $v_{i_1}^{(1)} \otimes ... \otimes v_{i_k}^{(k)}$.

So we define

$$v_{i_1}^{(1)} \otimes ... \otimes v_{i_k}^{(k)}(\tilde{i}_1, ..., \tilde{i}_k) := \begin{cases} 1 & \text{if } (\tilde{i}_1, ..., \tilde{i}_k) = (i_1, ..., i_k) \\ 0 & \text{else.} \end{cases}$$

$v_{i_1}^{(1)} \otimes ... \otimes v_{i_k}^{(k)}$ is the function that is 1 for $(i_1, ..., i_k)$ and 0 for all other arguments. We can form a basis of $V_1 \otimes ... \otimes V_k$ with $(v_{i_1}^{(1)} \otimes ... \otimes v_{i_k}^{(k)})_{(i_1,...,i_k) \in I_1 \times ... \times I_k}$ because:

- For any $\tau \in V_1 \otimes ... \otimes V_k$, we obtain

  $$\tau = \sum_{(i_1,...,i_k) \in I_1 \times ... \times I_k} \tau(i_1, ..., i_k)(v_{i_1}^{(1)} \otimes ... \otimes v_{i_k}^{(k)}).$$

  Hence $(v_{i_1}^{(1)} \otimes ... \otimes v_{i_k}^{(k)})_{(i_1,...,i_k) \in I_1 \times ... \times I_k}$ is a generator of $V_1 \otimes ... \otimes V_k$.

- Now the linear independence of these functions still needs to be shown.

  $$\sum_{(i_1,...,i_k) \in I_1 \times ... \times I_k} \alpha_{i_1,...,i_k}(v_{i_1}^{(1)} \otimes ... \otimes v_{i_k}^{(k)}) = 0$$
  $$\Rightarrow \forall (i_1, ..., i_k) \in I_1 \times ... \times I_k : \alpha_{i_1,...,i_k} = 0,$$

  because the constant 0-function maps every argument to 0 and all alone $\alpha_{i_1,...,i_k}$ sets the function value of the argument $(i_1, ..., i_k)$.

Now we define η by

$$\eta(v_{i_1}^{(1)}, ..., v_{i_k}^{(k)}) := v_{i_1}^{(1)} \otimes ... \otimes v_{i_k}^{(k)}.$$

There exists exactly one such η by Lemma 3.2.

The universal property of η remains to prove. Let W be a K-vector space and $\xi : V_1 \times ... \times V_k \to W$ a multilinear function with

$$w_{i_1...i_k} := \xi(v_{i_1}^{(1)}, ..., v_{i_k}^{(k)}).$$

Because $\xi = \xi_\otimes \circ \eta$ should apply, it has to be

$$\xi_\otimes(v_{i_1}^{(1)} \otimes ... \otimes v_{i_k}^{(k)}) = w_{i_1...i_k}.$$

Again we know by Lemma 3.2 that there exists exactly one such linear function $\xi_\otimes$. So we take this $\xi_\otimes$ and check if it satisfies the required conditions. For an arbitrary $v_1 \otimes ... \otimes v_k \in V_1 \otimes ... \otimes V_k$ with $v_j = \sum'_{i_j} \lambda_{j,i_j} v_{i_j}^{(j)}$, $j \in \{1, ..., k\}$, there is

$$\begin{aligned}
\xi_\otimes(v_1 \otimes ... \otimes v_k) &= \xi_\otimes \left(\sum_{(i_1,...,i_k) \in I_1 \times ... \times I_k}' \lambda_{1,i_1} ... \lambda_{k,i_k} (v_{i_1}^{(1)} \otimes ... \otimes v_{i_k}^{(k)}) \right) \\
&= \sum_{(i_1,...,i_k) \in I_1 \times ... \times I_k}' \lambda_{1,i_1} ... \lambda_{k,i_k} \xi_\otimes(v_{i_1}^{(1)} \otimes ... \otimes v_{i_k}^{(k)}) \\
&= \sum_{(i_1,...,i_k) \in I_1 \times ... \times I_k}' \lambda_{1,i_1} ... \lambda_{k,i_k} w_{i_1...i_k} \\
&= \sum_{(i_1,...,i_k) \in I_1 \times ... \times I_k}' \lambda_{1,i_1} ... \lambda_{k,i_k} \xi(v_{i_1}^{(1)}, ..., v_{i_k}^{(k)}) \\
&= \sum_{(i_1,...,i_k) \in I_1 \times ... \times I_k}' \xi(\lambda_{1,i_1} v_{i_1}^{(1)}, ..., \lambda_{k,i_k} v_{i_k}^{(k)}) \\
&= \sum_{i_1}' ... \sum_{i_k}' \xi(\lambda_{1,i_1} v_{i_1}^{(1)}, ..., \lambda_{k,i_k} v_{i_k}^{(k)}) \\
&= \xi \left(\sum_{i_1}' \lambda_{1,i_1} v_{i_1}^{(1)}, ..., \sum_{i_k}' \lambda_{k,i_k} v_{i_k}^{(k)} \right) \\
&= \xi(v_1, ..., v_k).
\end{aligned}$$

Thus, for every K-vector space W and every multilinear function $\xi : V_1 \times ... \times V_k \to W$ there exists a linear function $\xi_\otimes : V_1 \otimes ... \otimes V_k \to W$ such that $\xi = \xi_\otimes \circ \eta$.

Finally we assume that $V_1, ..., V_k$ are finite-dimensional. Then

$$\dim(V_1 \otimes ... \otimes V_k) = \#I_1 \cdot ... \cdot \#I_k = \dim V_1 \cdot ... \cdot \dim V_k.$$

□

After the introduction to the tensor product by the universal property, we may not have a concrete notion of what a tensor product can be. Note that the tensor product of vector spaces is unique up to isomorphism. So if we give an example of a tensor product here, this does not mean that the tensor product of these vector spaces must have this structure. We could also take an isomorphic vector space as tensor product.

Example. *We consider the vector spaces* $\mathbb{R}^m$ *and* $\mathbb{R}^n$*. We define*

$$\mathbb{R}^m \otimes \mathbb{R}^n := \mathbb{R}^{m\times n}$$

together with the multilinear function

$$\theta : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^{m\times n},\ \theta(x,y) := x \cdot y^T.$$

Given any vector space W *and an arbitrary multilinear function* $\xi : \mathbb{R}^m \times \mathbb{R}^n \to W$ *with* $(e_i, e_j) \mapsto w_{ij}$ *for all* $i \in \{1,...,m\}$ *and* $j \in \{1,...,n\}$*, we define a linear function* $\xi_\otimes : \mathbb{R}^{m\times n} \to W$ *by*

$$E_{ij} \mapsto w_{ij}$$

for all $i \in \{1,...,m\}$ *and* $j \in \{1,...,n\}$*, whereby* $E_{ij} \in \mathbb{R}^{m\times n}$ *denotes the matrix whose entries are all zero except the entry at position* (i,j)*, which is equal to* 1*. Indeed we have*

$$\begin{aligned}
\xi_\otimes(x \cdot y^T) &= \xi_\otimes\left(\sum_{\substack{1\le i\le m,\\ 1\le j\le n}} x_i y_j E_{ij}\right)\\
&= \sum_{\substack{1\le i\le m,\\ 1\le j\le n}} x_i y_j \xi_\otimes(E_{ij})\\
&= \sum_{\substack{1\le i\le m,\\ 1\le j\le n}} x_i y_j w_{ij}\\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} \xi(x_i e_i, y_j e_j)\\
&= \xi\left(\sum_{i=1}^{m} x_i e_i, \sum_{j=1}^{n} y_j e_j\right)\\
&= \xi(x,y).
\end{aligned}$$

So we actually defined a tensor product for the vector spaces $\mathbb{R}^n$ *and* $\mathbb{R}^m$*.*

3.4 Exterior product

After we have shown the existence of the tensor product, we will use it to prove that there is another product, the so-called *exterior product*. It becomes important if we want to consider multilinear functions that are additionally alternating. First we get to know the concept of *quotient spaces* and some statement about them. Again, this section's content is taken from Fischer's book [Fis13].

Definition 3.4 (quotient space)**.** Let V be a K-vector space and $U \subseteq V$ a subspace of V. The *quotient space* V/U is defined as the factor group V/U of the group $(V,+)$ and its subgroup $(U,+)$.

We cannot only define an addition on the quotient space. We are also able to define a scalar multiplication on it such that the quotient space becomes a vector space by itself.

Lemma 3.3. *Let V be a K-vector space and $U \subseteq V$ a subspace of V. Then we can assign a structure of a K-vector space to V/U such that the canonial function*

$$\rho : V \to V/U,\ \rho(v) := [v]_{\sim_U} = v + U$$

is linear. Furthermore, the following applies:

(i) *ρ is surjective.*

(ii) $\operatorname{Ker} \rho = U$.

(iii) $\dim V/U = \dim V - \dim U$ *if* $\dim V < \infty$.

(iv) *V/U has the following universal property: For every K-vector space W and linear function $F : V \to W$ with $U \subseteq \operatorname{Ker} F$, there exists a linear function $\bar{F} : V/U \to W$ such that $F = \bar{F} \circ \rho$. Moreover* $\operatorname{Ker} \bar{F} = (\operatorname{Ker} F)/U$.

Proof. First we try to assign a vector space structure to V/U. To do this, we need to find two operations: a vector addition $\dot{+}$ and a scalar multiplication $\cdot$. In the following, the old vector addition on V is denoted by $+$ and the scalar multiplication by no symbol. We want ρ to be linear. So we can already conclude that $\dot{+}$ and $\cdot$ have to fulfill

$$(v + U)\dot{+}(w + U) = \rho(v)\dot{+}\rho(w) = \rho(v + w) = (v + w) + U$$

and

$$\lambda \cdot (v + U) = \lambda \cdot \rho(v) = \rho(\lambda v) = \lambda v + U.$$

Hence, there is only one way to define $\dot{+}$ and $\cdot$, namely,

$$(v + U)\dot{+}(w + U) := (v + w) + U$$

and

$$\lambda \cdot (v + U) := \lambda v + U.$$

We already know by Lemma 2.1 that $\dot{+}$ is well-defined. The scalar multiplication is also well-defined. To verify this, assume arbitrary $v, v' \in V$ and $\lambda \in K$ such that $v + U = v' + U$. This means we have $v - v' \in U$, so $\lambda(v - v') = \lambda v - \lambda v' \in U$. Therefore $\lambda v + U = \lambda v' + U$, hence

$$\lambda \cdot (v + U) = \lambda v + U = \lambda v' + U = \lambda \cdot (v' + U).$$

The vector space axioms hold for $(V/U, \dot{+}, \cdot)$. Indeed we already know by Lemma 2.1 that $(V, \dot{+})$ is an abelian group. The remaining axioms follow directly by applying the vector space properties of V together with the definitions of $\dot{+}$ and $\cdot$. Note that we replace $\dot{+}$ and $\cdot$ by the common symbols for vector addition and scalar multiplication from now on.

We prove $(i), (ii), (iii)$ and (iv) now.

(i) For an arbitrary set $S \in V/U$, there exists $v \in V$ such that $S = v + U = \rho(v)$.

(*ii*) For $u \in U$ there is $\rho(u) = u + U = U$. Vice versa, if we have $\rho(v) = U$, we know that $v + U = U$, so v must be a vector in U.

(*iii*) It is well-known by the Rank-nullity theorem that $\dim V = \dim \operatorname{Im} \rho + \dim \operatorname{Ker} \rho$. From (*i*), we conclude that $\operatorname{Im} \rho = V/U$ and from (*ii*) that $\operatorname{Ker} \rho = U$. So $\dim V/U = \dim \operatorname{Im} \rho = \dim V - \dim \operatorname{Ker} \rho = \dim V - \dim U$.

(*iv*) Let $F : V \to W$ be a linear function with $U \subseteq \operatorname{Ker} F$. We want to define $\bar{F} : V/U \to W$ in such a way that $F = \bar{F} \circ \rho$. Hence for $v \in V$ holds

$$F(v) = \bar{F}(\rho(v)) = \bar{F}(v + U).$$

By defining $\bar{F}$ like this, that is, $\bar{F}(v + U) := F(v)$, we actually well-define $\bar{F}$ because: If $v + U = v' + U$ then $v - v' \in U \subseteq \operatorname{Ker} F$. So $F(v - v') = F(v) - F(v') = 0$, which means $F(v) = F(v')$. From

$$\begin{aligned}\bar{F}(\lambda(v + U) + (w + U)) = \bar{F}((\lambda v + w) + U) = F(\lambda v + w) = \lambda F(v) + F(w) \\ = \lambda \bar{F}(v + U) + \bar{F}(w + U)\end{aligned}$$

follows the linearity of $\bar{F}$. We prove the equation $\operatorname{Ker} \bar{F} = (\operatorname{Ker} F)/U$ by

$$v + U \in \operatorname{Ker} \bar{F} \Leftrightarrow \bar{F}(v + U) = 0 \Leftrightarrow F(v) = 0 \Leftrightarrow v + U \in (\operatorname{Ker} F)/U.$$

□

There are special linear functions, the so-called alternating functions. In the following, V^k denotes the k-fold Cartesian product $V \times ... \times V$ of V.

Definition 3.5 (alternating functions)**.** Let V, W be K-vector spaces and $\xi : V^k \to W$ a multilinear function. ξ is called *alternating* if for all $v_1, ..., v_k \in V$

$$\xi(v_1, ..., v_k) = 0$$

applies if $v_i = v_j$ for a pair (i, j) with $i \neq j$.

We could give the determinant function defined on the n-fold Cartesian product $K^n \times ... \times K^n$ instead of $K^{n \times n}$ as an example. It is well-known that this function is multilinear and alternating. Why such functions are actually called alternating will be clear by the next lemma.

Let $\sigma : \{1, ..., k\} \to \{1, ..., k\}$ be a permutation, then we define the **number of inversions** $N(\sigma)$ of σ as the cardinality of the set

$$\{(i, j) \in \{1, ..., k\} \times \{1, ..., k\} \mid i < j, \sigma(i) > \sigma(j)\}.$$

Moreover, the **signum** $\operatorname{sign}(\sigma)$ of σ is defined as $(-1)^{N(\sigma)}$.

Lemma 3.4. *Let $\xi : V^k \to W$ be an alternating function, $v_1, ..., v_k \in V$ and $\sigma : \{1, ..., k\} \to \{1, ..., k\}$ a permutation. Then*

$$\xi(v_1, ..., v_k) = \operatorname{sign}(\sigma) \xi(v_{\sigma(1)}, ..., v_{\sigma(k)}).$$

Proof. Let $i, j \in \{1, ..., k\}, i < j$. Then

$$\begin{aligned} 0 &= \xi(v_1, ..., \underbrace{v_i + v_j}_{i\text{-th position}}, ..., \underbrace{v_i + v_j}_{j\text{-th position}}, ..., v_k) \\ &= \xi(v_1, ..., v_i, ..., v_j, ..., v_k) + \underbrace{\xi(v_1, ..., v_i, ..., v_i, ..., v_k)}_{=0} \\ &+ \underbrace{\xi(v_1, ..., v_j, ..., v_j, ..., v_k)}_{=0} + \xi(v_1, ..., v_j, ..., v_i, ..., v_k), \end{aligned}$$

hence

$$\xi(v_1, ..., v_i, ..., v_j, ..., v_k) = (-1)\xi(v_1, ..., v_j, ..., v_i, ..., v_k).$$

Let $N(\sigma)$ be the number of transpositions of σ. Then $(v_{\sigma(1)}, ..., v_{\sigma(k)})$ is obtained by applying $N(\sigma)$ transpositions on $(v_1, ..., v_k)$. Therefore we get

$$\begin{aligned} \xi(v_1, ..., v_k) &= \underbrace{(-1) \cdot ... \cdot (-1)}_{N(\sigma) \text{ times}} \xi(v_{\sigma(1)}, ..., v_{\sigma(k)}) \\ &= (-1)^{N(\sigma)} \xi(v_{\sigma(1)}, ..., v_{\sigma(k)}) \\ &= \operatorname{sign}(\sigma) \xi(v_{\sigma(1)}, ..., v_{\sigma(k)}). \end{aligned}$$

□

So whenever we exchange two arguments of our function, the sign of the function value will change. That is, the function alternates when we exchange arguments.

If we have a vector space V over K and a subset S of V, then we define the **linear span** $\operatorname{span}_K S$ as the intersection of all subspaces of V that contain S. It is well-known that

$$\operatorname{span}_K S = \left\{ \sum_{i=1}^{n} \lambda_i v_i \;\middle|\; n \in \mathbb{N}, \lambda_i \in K, v_i \in S \text{ for all } 1 \leq i \leq n \right\}.$$

If the field K is clear from the context, we just want to write $\operatorname{span} S$ instead of $\operatorname{span}_K S$.

Definition 3.6. Let $\bigotimes^k V := V \otimes ... \otimes V$ be the k-fold tensor product of V. Then we define

$$A^k(V) := \operatorname{span}\{v_1 \otimes ... \otimes v_k\}_{\exists i,j \in \{1,...,k\}, i \neq j : v_i = v_j}.$$

The reason why we are defining this is that we want to divide the tensor product by this subspace to obtain the exterior product later on.

Lemma 3.5. *For arbitrary multilinear $\xi : V^k \to W$*

$$\xi \text{ alternating } \Leftrightarrow A^k(V) \subseteq \operatorname{Ker} \xi_{\otimes}.$$

Proof. We just need to prove both implications.

"⇒" Assume that ξ is alternating and let $v \in A^k(V)$. Then there exist $n \in \mathbb{N}$, $\lambda_1, ..., \lambda_n \in K$ and $w_1, ..., w_k \in \{v_1 \otimes ... \otimes v_k\}_{\exists i,j \in \{1,...,k\}, i \neq j : v_i = v_j}$ such that

$$v = \sum_{j=1}^{n} \lambda_j w_j.$$

We know that every w_t, $t \in \{1, ..., n\}$, has the form $v_1^{(t)} \otimes ... \otimes v_k^{(t)}$ for $v_1^{(t)}, ..., v_k^{(t)} \in V$ with $v_i^{(t)} = v_j^{(t)}$ for some $i \neq j$. We obtain

$$\begin{aligned}\xi_\otimes(v) = \xi_\otimes\left(\sum_{j=1}^{n} \lambda_j w_j\right) = \sum_{j=1}^{n} \lambda_j \xi_\otimes(w_j) = \sum_{j=1}^{n} \lambda_j \xi_\otimes(v_1^{(j)} \otimes ... \otimes v_k^{(j)}) \\ = \sum_{j=1}^{n} \lambda_j \xi(v_1^{(j)}, ..., v_k^{(j)}) = \sum_{j=1}^{n} \lambda_j 0 = 0.\end{aligned}$$

"⇐" This direction is trivial since

$$\xi_\otimes(v_1 \otimes ... \otimes v_k) = \xi(v_1, ..., v_k).$$

□

Finally, we can state the theorem which verifies the existence of the exterior product.

Theorem 3. *Let V be a K-vector space. Then there exist a K-vector space $\bigwedge^k V$ and an alternating function*

$$\wedge : V^k \to \bigwedge^k V,\ \wedge(v_1, ..., v_k) := v_1 \wedge ... \wedge v_k$$

that have the following universal property: For all K-vector spaces W together with an alternating function

$$\xi : V^k \to W,$$

there exists exactly one linear function

$$\xi_\wedge : \bigwedge^k V \to W$$

such that $\xi = \xi_\wedge \circ \wedge$. If $(v_1, ..., v_n)$ is a basis of V, we obtain a basis of $\bigwedge^k V$ by

$$(v_{i_1} \wedge ... \wedge v_{i_k})_{1 \leq i_1 < ... < i_k \leq n}.$$

In particular,

$$\dim \bigwedge^k V = \binom{n}{k}$$

for $1 \leq k \leq n = \dim V$.

Remark. We set $\bigwedge^0 V := K$ and $\bigwedge^k V := \{0\}$ for $k > n$.

Proof. We define

$$\bigwedge^k V := \bigotimes^k V / A^k(V).$$

Let $\rho : \bigotimes^k V \to \bigotimes^k V/A^k(V)$ be the canonial linear function. Then we define

$$\wedge := \rho \circ \eta,$$

whereby η is taken from Theorem 2. For $v_1, ..., v_k \in V$, we obtain

$$v_1 \wedge ... \wedge v_k := \wedge(v_1, ..., v_k) = \rho(\eta(v_1, ..., v_k)) = \rho(v_1 \otimes ... \otimes v_k).$$

η is multilinear and ρ is linear. Hence $\wedge$ is multilinear as composition of this two functions. Because

$$\operatorname{Ker} \rho = A^k(V),$$

which follows from Lemma 3.3, we know by Lemma 3.5 that $\wedge$ is alternating. (To unterstand this, we could informally write down "$\rho = \wedge_\otimes$".)

The universal property remains to be proven. So let $\xi : V^k \to W$ be an alternating function. We know by Theorem 2 that there exists a unique linear $\xi_\otimes : \bigotimes^k V \to W$ with $\xi = \xi_\otimes \circ \eta$. Because ξ is alternating, we know that

$$A^k(V) \subseteq \operatorname{Ker} \xi_\otimes$$

and therefore, due to the universal property of the quotient space, we obtain the unique existence of a linear function

$$\xi_\wedge : \underbrace{\bigotimes^k V / A^k(V)}_{= \bigwedge^k V} \to W$$

with $\xi_\otimes = \xi_\wedge \circ \rho$. This implies $\xi_\otimes \circ \eta = \xi_\wedge \circ \rho \circ \eta$, in fact

$$\xi = \xi_\wedge \circ \rho \circ \eta = \xi_\wedge \circ \wedge.$$

At last we show the statements about the dimension of the exterior product: Let $(v_1, ..., v_n)$ be a basis of V. We know that $(v_{i_1} \otimes ... \otimes v_{i_k})_{1 \le i_1, ..., i_k \le n}$ is a basis of the k-fold tensor product $\bigotimes^k V$ of V. $\wedge$ is a multilinear function, hence we know that

$$\operatorname{Im} \wedge = \operatorname{span}\{\wedge(v_{i_1}, ..., v_{i_k})\}_{1 \le i_1, ..., i_k \le n} = \operatorname{span}\{v_{i_1} \wedge ... \wedge v_{i_k}\}_{1 \le i_1, ..., i_k \le n}.$$

In fact, $G := (v_{i_1} \wedge ... \wedge v_{i_k})_{1 \le i_1, ..., i_k \le n}$ is a generator of $\bigwedge^k V$. Because $\wedge$ is alternating,

$$(\exists s, t \in \{1, ..., n\} : v_{i_s} = v_{i_t}) \Rightarrow v_{i_1} \wedge ... \wedge v_{i_k} = 0.$$

Therefore we can delete those vectors from G and it is still a generator. Moreover we know by Lemma 3.4 that exchanging the arguments of $\wedge$ only changes the sign of the image. So for $w_1, ..., w_k \in V$ we know that $(w_{\sigma(1)} \wedge ... \wedge w_{\sigma(k)})_{\sigma \text{ permutation of } 1,...,k}$ is a linearly dependent family of vectors. That is why it is enough to keep only vectors $v_{i_1} \wedge ... \wedge v_{i_k}$ in G with sorted indices $i_1 < ... < i_k$. Summarized we can state that

$$(v_{i_1} \wedge ... \wedge v_{i_k})_{1 \le i_1 < ... < i_k \le n}$$

is a generator for $\bigwedge^k V$ containing $\binom{n}{k}$ vectors. The linear independence of those vectors remains to be shown. We do this by creating an isomorphism to K^N, whereby $N := \binom{n}{k}$. Let $(e_{i_1 \dots i_k})_{1 \le i_1 < \dots < i_k \le n}$ be the canonial basis of K^N, $w_1, \dots, w_k \in V$ with

$$\forall i \in \{1, \dots, k\} : w_i = \sum_{j=1}^{n} \lambda_{ij} v_j$$

and $A := (\lambda_{ij})_{i=1,\dots,k, j=1,\dots,n}$. Further, let $a_{i_1 \dots i_k}$ be the minor of A belonging to the columns $i_1, \dots, i_k$. That is,

$$a_{i_1 \dots i_k} := \det(\lambda_{ij})_{i=1,\dots,k, j=i_1,\dots,i_k}.$$

Now we define

$$\xi : V^k \to K^N, \xi(w_1, \dots, w_k) := \sum_{\substack{i_1,\dots,i_k \text{ with} \\ 1 \le i_1 < \dots < i_k \le n}} a_{i_1 \dots i_k} e_{i_1 \dots i_k}.$$

ξ is alternating. To verify this, let $w_s = w_t$ for some $s \neq t$. Then w_s and w_t are built by the same linear combination of the basis vectors $v_1, \dots, v_k$. So two rows in A are equal. Thus, every $a_{i_1 \dots i_k}$ is zero since the determinant function is alternating. We obtain $\xi(w_1, \dots, w_k) = 0$. Now we know that there exists $\xi_\wedge : \bigwedge^k V \to K^N$ because of the universal property with

$$\xi_\wedge(v_{i_1} \wedge \dots \wedge v_{i_k}) = \xi(v_{i_1}, \dots, v_{i_k}) = \sum_{\substack{i_1,\dots,i_k \text{ with} \\ 1 \le i_1 < \dots < i_k \le n}} a_{i_1 \dots i_k} e_{i_1 \dots i_k} \overset{(!)}{=} e_{i_1 \dots i_k},$$

whereby $1 \le i_1 < \dots < i_k \le n$. To unterstand the rightmost equality, we have to look at the corresponding matrix A. We know that $v_{i_l} = \sum_{j=1}^{n} \delta_{i_l j} v_j$ for all $l \in \{1, \dots, k\}$. Therefore,

$$A = (\delta_{i_l j})_{l=1,\dots,k, j=1,\dots,n}.$$

That is, the l-th row has a one-entry in column position i_l. Because all i_l are different, we have exactly k columns that have an entry unequal zero, namely, the i_1-th, i_2-th, ... and i_k-th column. So only $a_{i_1 \dots i_k} = 1 \neq 0$ because all the minors belonging to other combinations of columns of A are the determinant of a matrix containing a zero-column. Finally, from $\xi_\wedge(v_{i_1} \wedge \dots \wedge v_{i_k}) = e_{i_1 \dots i_k}$ and the linear independence of $(e_{i_1 \dots i_k})_{1 \le i_1 < \dots < i_k \le n}$ follows the linear independence of $(v_{i_1} \wedge \dots \wedge v_{i_k})_{1 \le i_1 < \dots < i_k \le n}$. Hence, this family is a basis of $\bigwedge^k V$ and $\dim \bigwedge^k V = \binom{n}{k}$. □

Remark. Note that $\bigwedge^k V$ and K^N are isomorphic vector spaces.

To make this notion clear, we discuss a straightforward example.

Example. *We want to prove that the cross product of two three-dimensional vectors defines an exterior product. So we define* $\wedge : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3 \wedge \mathbb{R}^3 := \mathbb{R}^3$ *by*

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \wedge \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} := \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix}.$$

We can obviously see that this function is multilinear and even alternating. Every alternating function $\xi : \mathbb{R}^3 \times \mathbb{R}^3 \to W$, whereby W is any vector space, can be translated uniquely into a linear function $\xi_\wedge : \mathbb{R}^3 \to W$ by defining

$$\xi_\wedge(e_1) := \xi(e_2, e_3),\ \xi_\wedge(e_2) := -\xi(e_1, e_3),\ \xi_\wedge(e_3) := \xi(e_1, e_2).$$

Indeed, we obtain for all $a = (a_1, a_2, a_3)^T, b = (b_1, b_2, b_3)^T \in \mathbb{R}^3$

$$\begin{aligned}
\xi_\wedge(a \wedge b) &= (a_2b_3 - a_3b_2)\xi_\wedge(e_1) + (a_3b_1 - a_1b_3)\xi_\wedge(e_2) + (a_1b_2 - a_2b_1)\xi_\wedge(e_3) \\
&= (a_2b_3 - a_3b_2)\xi(e_2, e_3) + (a_3b_1 - a_1b_3)(-\xi(e_1, e_3)) + (a_1b_2 - a_2b_1)\xi(e_1, e_2) \\
&= a_1b_2\xi(e_1, e_2) + a_1b_3\xi(e_1, e_3) - a_2b_1\xi(e_1, e_2) + a_2b_3\xi(e_2, e_3) \\
&\quad - a_3b_1\xi(e_1, e_3) - a_3b_2\xi(e_2, e_3) \\
&= a_1b_2\xi(e_1, e_2) + a_1b_3\xi(e_1, e_3) + a_2b_1\xi(e_2, e_1) + a_2b_3\xi(e_2, e_3) \\
&\quad + a_3b_1\xi(e_3, e_1) + a_3b_2\xi(e_3, e_2) \\
&= a_1b_1\xi(e_1, e_1) + a_1b_2\xi(e_1, e_2) + a_1b_3\xi(e_1, e_3) + a_2b_1\xi(e_2, e_1) + a_2b_2\xi(e_2, e_2) \\
&\quad + a_2b_3\xi(e_2, e_3) + a_3b_1\xi(e_3, e_1) + a_3b_2\xi(e_3, e_2) + a_3b_3\xi(e_3, e_3) \\
&= \xi(a_1e_1 + a_2e_2 + a_3e_3, b_1e_1 + b_2e_2 + b_3e_3) \\
&= \xi(a, b).
\end{aligned}$$

Hence, the cross product is an exterior product. Also we want to verify the dimension of the exterior product. Indeed we have $n = 3$ and $k = 2$, whereby n and k are the notations from the above theorem. Hence, $\binom{3}{2} = 3$, which is equal to the dimension of $\mathbb{R}^3 \wedge \mathbb{R}^3 = \mathbb{R}^3$.

There are some helpful properties of the exterior product [Reu80].

Lemma 3.6. *Let $v, w \in V$. Then*

$$v \wedge w = -(w \wedge v).$$

Proof. $\wedge$ is alternating, so we can apply Lemma 3.4. □

Lemma 3.7. *Let $v_1, ..., v_k \in V$. Then*

$$v_1, ..., v_k \text{ linearly dependent} \Leftrightarrow v_1 \wedge ... \wedge v_k = 0.$$

Proof. We conclude in two steps.

"⇒" If $v_1, ..., v_k$ are linearly dependent, we know that there exist $i \in \{1, ..., k\}$ and $\lambda_1, ..., \lambda_{i-1}, \lambda_{i+1}, ..., \lambda_k \in K$ such that

$$v_i = \sum_{j \in \{1,...,k\} \setminus \{i\}} \lambda_j v_j.$$

Then

$$\begin{aligned}
v_1 \wedge ... \wedge v_i \wedge ... \wedge v_k &= v_1 \wedge ... \wedge \left(\sum_{j \in \{1,...,k\} \setminus \{i\}} \lambda_j v_j \right) \wedge ... \wedge v_k \\
&= \sum_{j \in \{1,...,k\} \setminus \{i\}} \lambda_j \underbrace{(v_1 \wedge ... \wedge \overbrace{v_j}^{i\text{-th position}} \wedge ... \wedge v_k)}_{=0} \\
&= 0.
\end{aligned}$$

"⇐" Let $v_1, ..., v_k \in V$ be linearly independent. We can extend these vectors to a basis $(v_i)_{i \in I}$ of V with $1, ..., k \in I$. Further, let $w_1, ..., w_k \in V$ with

$$\forall i \in \{1, ..., k\} : w_i = \sum_{j \in I} \lambda_{ij} v_j.$$

This leads to the definition of

$$\xi : V^k \to K,\ \xi(w_1, ..., w_k) := \det(\lambda_{ij})_{i,j=1,...,k},$$

which is an alternating function. So there exists a linear $\xi_\wedge$ with $\xi = \xi_\wedge \circ \wedge$. We can state

$$\xi_\wedge(v_1 \wedge ... \wedge v_k) = \xi(v_1, ..., v_k) = \det(\mathbb{I}_k) = 1 \neq 0.$$

Hence, $v_1 \wedge ... \wedge v_k \neq 0$.

□

Definition 3.7. Let V be a K-vector space, $U \subseteq V$ a subspace of V and $(b_1, ..., b_n)$ a basis of U. Then we define

$$\bar{U} := b_1 \wedge ... \wedge b_n.$$

In case $U = \{0\}$, we set $\bar{U} := 1$.

Remark. $\bar{U}$ is not well-defined. But the following lemma states that we can still use the definition if a constant factor does not matter.

Lemma 3.8. *Let $(b_1, ..., b_n)$ and $(b'_1, ..., b'_n)$ be bases of U. Then there exists $c \in K, c \neq 0$ such that*

$$b'_1 \wedge ... \wedge b'_n = c \cdot (b_1 \wedge ... \wedge b_n).$$

That is, $\bar{U}$ is unique except for constants.

Proof. There exist $\lambda_{ij} \in K, i, j \in \{1, ..., n\}$ such that

$$\begin{aligned}
b'_1 \wedge ... \wedge b'_n &= \left(\sum_{j=1}^n \lambda_{1j} b_j\right) \wedge ... \wedge \left(\sum_{j=1}^n \lambda_{nj} b_j\right) \\
&= \sum_{j_1=1}^n ... \sum_{j_n=1}^n \lambda_{1j_1} ... \lambda_{nj_n} \cdot (b_{j_1} \wedge ... \wedge b_{j_n}) \\
&\overset{\text{Lemma 3.4}}{=} \sum_{j_1=1}^n ... \sum_{j_n=1}^n \lambda_{1j_1} ... \lambda_{nj_n} \operatorname{sign}(j_1, ..., j_n) \cdot (b_1 \wedge ... \wedge b_n) \\
&= (b_1 \wedge ... \wedge b_n) \cdot \underbrace{\sum_{j_1=1}^n ... \sum_{j_n=1}^n \lambda_{1j_1} ... \lambda_{nj_n} \operatorname{sign}(j_1, ..., j_n)}_{=:c}.
\end{aligned}$$

$c \neq 0$ because c is the determinant of the transformation matrix of the identity function with bases $(b_1, ..., b_n)$ and $(b'_1, ..., b'_n)$, which is bijective. □

Lemma 3.9. *Let $E, F \subseteq V$ be subspaces of V. Then*

$$E \cap F = \{0\} \Leftrightarrow \bar{E} \wedge \bar{F} \neq 0.$$

Proof. Again we prove both directions.

"$\Rightarrow$" For arbitrary bases $(e_1, ..., e_k)$ of E and $(f_1, ..., f_l)$ of F, we know due to the fact that E and F are disjoint that $(e_1, ..., e_k, f_1, ..., f_l)$ is a linearly independent family of vectors. Hence, $\bar{E} \wedge \bar{F} \neq 0$.

"$\Leftarrow$" The assumption implies that every basis of E combined with an arbitrary basis of F results in a linearly independent system of vectors. Therefore E and F must be disjoint subspaces.

□

At last note that

$$\bar{U} \in \bigwedge^{\dim(U)} V.$$

3.5 The existence of a pseudo-regular subproduct

The following theorem claims that we can find a pseudo-regular matrix in every matrix-product which is long enough.

Theorem 4 (Reutenauer [Reu80])**.** *Let $n \in \mathbb{N}$. There exists $N \in \mathbb{N}$ such that for all $(A_i)_{1 \leq i \leq N} \in (K^{n \times n})^{\{1,...,N\}}$ there exist $1 \leq i \leq j \leq N$ such that $A_i...A_j$ is a pseudo-regular matrix.*

Proof. At first we define $k_0 := 1, k_n := 1$ and $k_i := \binom{n}{i} + 1$ for all $i \in \{1, ..., n-1\}$. Let $N := k_0...k_n$ and $A := (A_i)_{1 \leq i \leq N} \in (K^{n \times n})^{\{1,...,N\}}$ be an arbitrary family of N square matrices. We know by Lemma 3.1 that there exists $l \in \{0, ..., n\}$ and a factor of the form $(B_1, ..., B_{k_l})$ with

$$\forall 1 \leq i \leq j \leq k_l : \operatorname{rank}(B_i...B_j) = l.$$

We conclude by considering three cases.

If $l = 0$, $\operatorname{rank}(B_1) = 0$, that means B_1 is the zero matrix hence pseudo-regular. Because $B_1 = A_i...A_j$ for some $1 \leq i \leq j \leq N$, we are done.

If $l = n$, $\operatorname{rank}(B_1) = n$. Therefore B_1 is invertible hence pseudo-regular. Again, we are done.

The remaining case is $1 \leq l \leq n-1$. We define

$$E_i := \operatorname{Im}(B_i), F_i := \operatorname{Ker}(B_i)$$

for every $1 \leq i \leq k_l$. Furthermore, we know for all $j \in \{1, ..., k_l - 1\}$ that

$$\operatorname{rank}(B_j B_{j+1}) = \operatorname{rank}(B_{j+1}).$$

From that and $\mathrm{Ker}(B_{j+1}) \subseteq \mathrm{Ker}(B_j B_{j+1})$ follows the equality of the kernels. Let $v \in \mathrm{Im}(B_{j+1}) \cap \mathrm{Ker}(B_j)$, then $B_j v = 0$ and it exists $w \in K^{n\times 1}$ such that $B_{j+1} w = v$. By substitution, we obtain $B_j B_{j+1} w = 0$, which implies $B_{j+1} w = 0 = v$. Hence

$$\mathrm{Im}(B_{j+1}) \cap \mathrm{Ker}(B_j) = \{0\} \Rightarrow E_{j+1} \cap F_j = \{0\} \Rightarrow \underbrace{\bar{E}_{j+1} \wedge \bar{F}_j \neq 0}_{(*)}.$$

But we can also state for all $1 \leq i \leq j \leq k_l$ that

$$\begin{aligned} &\mathrm{rank}(B_i ... B_j) = \mathrm{rank}(B_i) = \mathrm{rank}(B_j), \\ &\mathrm{Im}(B_i ... B_j) \subseteq \mathrm{Im}(B_i) \text{ and} \\ &\mathrm{Ker}(B_j) \subseteq \mathrm{Ker}(B_i ... B_j), \end{aligned}$$

hence

$$\mathrm{Im}(B_i ... B_j) = \mathrm{Im}(B_i) = E_i \text{ and } \mathrm{Ker}(B_i ... B_j) = \mathrm{Ker}(B_j) = F_j.$$

We try to conclude by contradiction. Therefore assume that $B_i ... B_j$ is not pseudo-regular for all $1 \leq i \leq j \leq k_l$. Then we know by the definition of pseudo-regularity that

$$E_i \cap F_j = \mathrm{Im}(B_i ... B_j) \cap \mathrm{Ker}(B_i ... B_j) \neq \{0\}$$

and therefore

$$\bar{E}_i \wedge \bar{F}_j = 0. \tag{3}$$

We know for arbitrary $i \in \{1, ..., k_l\}$ that

$$\dim E_i = \dim \mathrm{Im}(B_i) = \mathrm{rank}(B_i) = l,$$

which implies that

$$\bar{E}_i \in \bigwedge^l K^n,$$

whereby the dimension of this vector space is $\binom{n}{l} = k_l - 1$. Therefore there exists $j \in \{1, ..., k_l - 1\}$ such that $\bar{E}_{j+1}$ is a linear combination of $\bar{E}_1, ..., \bar{E}_j$, that is,

$$\bar{E}_{j+1} = \sum_{i=1}^{j} \lambda_i \bar{E}_i$$

for some $\lambda_1, ..., \lambda_j \in K$. To verify this, first assume the worst case: All $\bar{E}_i, 1 \leq i \leq k_l - 1$, are linearly independent. But then they form a basis due to the dimension of the exterior product and E_{k_l} can be written as a linear combination. In the other case we get the desired linear combination anyway. Now we can state that

$$\bar{E}_{j+1} \wedge \bar{F}_j = \left(\sum_{i=1}^{j} \lambda_i \bar{E}_i \right) \wedge \bar{F}_j = \sum_{i=1}^{j} \lambda_i \underbrace{(\bar{E}_i \wedge \bar{F}_j)}_{=0 \text{ by } (3)} = 0,$$

which is a contradiction to $(*)$. So our assumption is false and we can conclude the existence of a pseudo-regular factor. □

n	N
1	1
2	3
3	16
4	175
5	4,356
6	263,424
7	40,144,896
8	15,714,084,159
9	15,953,234,222,500
10	42,223,789,335,548,788

Table 1: Product length depending on the dimension

We want to calculate some of these N's manually. Increasing the dimension of the matrices requires the product to be significantly longer. The above proof presented a way to calculate N depending on the matrix dimension n:

$$N(n) := \prod_{i=1}^{n-1} \left(\binom{n}{i} + 1 \right).$$

How the product length increases can be seen in Table 1.

Example. *We verify the result for* 2×2 *matrices by hand. Therefore, consider the following product:*

$$\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}.$$

All three matrices are not pseudo-regular. We can explain this by showing that the kernel and the image of $A := \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}$ *are not complementary subspaces. We have*

$$\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x - y \\ x - y \end{pmatrix}$$

for all $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$, *hence* $\operatorname{Im} A = \operatorname{Ker} A = \left\{ \begin{pmatrix} x \\ x \end{pmatrix} \in \mathbb{R}^2 \mid x \in \mathbb{R} \right\}$, *which means that these two subspaces are not complementary.*

But we know there has to exist a pseudo-regular subproduct because our product has a length greater or equal to three. Indeed this is actually true because AAA *and even* AA *are pseudo-regular since*

$$AA = AAA = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

3.6 Generalization to matrices over integral domains

We want to generalize the previous theorem to matrices with entries taken from an integral domain now. Therefore we first need a definition for pseudo-regular

matrices over integral domains that makes sense with respect to the final theorem of this thesis, which will be presented in the next chapter. Let R be an integral domain in the following.

Let $\mathbb{E}$ be the embedding $r \mapsto \frac{r}{1}$ of R in $\mathrm{Frac}(R)$. Then we want to let $\mathbb{E}$ act on a matrix $A = (a_{ij})_{\substack{1\le i\le m,\\ 1\le j\le n}} \in R^{m\times n}$ like

$$\mathbb{E}(A) := (\mathbb{E}(a_{ij}))_{\substack{1\le i\le m,,\\ 1\le j\le n}},$$

that is, we translate each entry of A, which is an element of R, into an entry taken from the field of fractions of R. We know that $\mathbb{E}$ is multiplicative and additive on matrices. Indeed, let $A = (a_{ij})_{\substack{1\le i\le l,\\ 1\le j\le m}} \in R^{l\times m}, B = (b_{ij})_{\substack{1\le i\le l,\\ 1\le j\le m}} \in R^{l\times m}$ and $C = (c_{ij})_{\substack{1\le i\le m,\\ 1\le j\le n}} \in R^{m\times n}$, then

$$\begin{aligned}
\mathbb{E}(A+B) &= \mathbb{E}\left((a_{ij}+b_{ij})_{\substack{1\le i\le l,\\ 1\le j\le m}}\right)\\
&= (\mathbb{E}(a_{ij}+b_{ij}))_{\substack{1\le i\le l,\\ 1\le j\le m}}\\
&= (\mathbb{E}(a_{ij})+\mathbb{E}(b_{ij}))_{\substack{1\le i\le l,\\ 1\le j\le m}}\\
&= (\mathbb{E}(a_{ij}))_{\substack{1\le i\le l,\\ 1\le j\le m}} + (\mathbb{E}(b_{ij}))_{\substack{1\le i\le l,\\ 1\le j\le m}}\\
&= \mathbb{E}(A)+\mathbb{E}(B)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}(A\cdot C) &= \mathbb{E}\left(\left(\sum_{k=1}^{m} a_{ik}c_{kj}\right)_{\substack{1\le i\le l,\\ 1\le j\le n}}\right)\\
&= \left(\mathbb{E}\left(\sum_{k=1}^{m} a_{ik}c_{kj}\right)\right)_{\substack{1\le i\le l,\\ 1\le j\le n}}\\
&= \left(\sum_{k=1}^{m} \mathbb{E}(a_{ik}c_{kj})\right)_{\substack{1\le i\le l,\\ 1\le j\le n}}\\
&= \left(\sum_{k=1}^{m} \mathbb{E}(a_{ik})\mathbb{E}(c_{kj})\right)_{\substack{1\le i\le l,\\ 1\le j\le n}}\\
&= (\mathbb{E}(a_{ij}))_{\substack{1\le i\le l,\\ 1\le j\le m}} \cdot (\mathbb{E}(c_{ij}))_{\substack{1\le i\le m,\\ 1\le j\le n}}\\
&= \mathbb{E}(A)\cdot\mathbb{E}(C),
\end{aligned}$$

because $\mathbb{E}$ is additive and multiplicative on R.

Definition 3.8 (pseudo-regular matrix over an integral domain)**.** Let $A \in R^{n\times n}$ and $\mathbb{E}$ the embedding of R in $\mathrm{Frac}(R)$. Then A is called *pseudo-regular* if $\mathbb{E}(A)$ is pseudo-regular by Definition 3.1, whereby we have $K = \mathrm{Frac}(R)$.

Remark. Note that $\mathbb{E}(A) \in \mathrm{Frac}(R)^{n\times n}$, that is, $\mathbb{E}(A)$ is a matrix with entries taken from a field. Also note that if R is a field, this definition is equivalent to Definition 3.1. Hence, we extended the definition for pseudo-regularity in a natural way.

The next lemma is already proved over fields by Reutenauer [Reu80]. We want to state it over integral domains and make use of the fact that the embedding of R in its field of fractions only maps the zero matrix to the zero matrix. This means that the kernel of the embedding, which is a ring homomorphism, is trivial. Note that this lemma is not needed to generalize the previous theorem. Rather, it is really important for the next chapter, in which we will prove the generalized iteration theorem for recognizable formal power series on trees.

Lemma 3.10. *Let $A \in R^{n\times n}, \lambda \in R^{1\times n}, \gamma \in R^{n\times 1}$ and $(p_k)_{k\in\mathbb{N}}$ a sequence defined by $p_k := \lambda A^k \gamma$. If A is pseudo-regular and $p_1 \neq 0$, the set $\{k \in \mathbb{N} \mid p_k \neq 0\}$ is infinite.*

Proof. Let P be the minimal polynomial of $\mathbb{E}(A)$ with $P(M) = \sum_{i=0}^{m} \alpha_{m-i} M^i \overset{\alpha_0=1}{=} M^m + \alpha_1 M^{m-1} + ... + \alpha_{m-1} M + \alpha_m M^0$, $m \in \mathbb{N}$ and $\alpha_i \in \mathrm{Frac}(R)$ for all $0 \leq i \leq m$.

Because A is pseudo-regular, we know from Proposition 3.1 that $\alpha_m \neq 0$ or $\alpha_{m-1} \neq 0$. Otherwise M^2 would divide $P(M)$.

Since $\mathbb{E}(A^k)P(\mathbb{E}(A)) = \mathbb{E}(A)^k P(\mathbb{E}(A)) = \mathbb{O}_{n\times n}$ for all $k \in \mathbb{N}$ by the multiplicativity of $\mathbb{E}$ and the property of the minimal polynomial, we define

$$\tilde{p}_k := \mathbb{E}(p_k) = \mathbb{E}(\lambda A^k \gamma)$$

and obtain

$$\begin{aligned}
0 &= \mathbb{E}(\lambda)\mathbb{E}(A)^k P(\mathbb{E}(A))\mathbb{E}(\gamma) \\
&= \mathbb{E}(\lambda)(\mathbb{E}(A)^{k+m} + \alpha_1 \mathbb{E}(A)^{k+m-1} + ... + \alpha_{m-1}\mathbb{E}(A)^{k+1} + \alpha_m \mathbb{E}(A)^k)\mathbb{E}(\gamma) \\
&= \mathbb{E}(\lambda)\mathbb{E}(A)^{k+m}\mathbb{E}(\gamma) + \alpha_1 \mathbb{E}(\lambda)\mathbb{E}(A)^{k+m-1}\mathbb{E}(\gamma) + ... \\
&\quad + \alpha_{m-1}\mathbb{E}(\lambda)\mathbb{E}(A)^{k+1}\mathbb{E}(\gamma) + \alpha_m \mathbb{E}(\lambda)\mathbb{E}(A)^k \mathbb{E}(\gamma) \\
&= \mathbb{E}(\lambda A^{k+m}\gamma) + \alpha_1 \mathbb{E}(\lambda A^{k+m-1}\gamma) + ... + \alpha_{m-1}\mathbb{E}(\lambda A^{k+1}\gamma) + \alpha_m \mathbb{E}(\lambda A^k \gamma) \\
&= \tilde{p}_{k+m} + \alpha_1 \tilde{p}_{k+m-1} + ... + \alpha_{m-1}\tilde{p}_{k+1} + \alpha_m \tilde{p}_k.
\end{aligned} \tag{4}$$

Let $i \in \mathbb{N}_+$ such that $p_i \neq 0$. Now we prove that there exists $j > i$ with $p_j \neq 0$. $p_i \neq 0$ implies $p_i \cdot 1 \neq 1 \cdot 0$ and therefore $\tilde{p}_i = \mathbb{E}(p_i) = \frac{p_i}{1} \neq \frac{0}{1} = 0$. We need to consider two cases now.

- If $\alpha_m \neq 0$, we know from equation (4) that

  $$\tilde{p}_{i+m} + \alpha_1 \tilde{p}_{i+m-1} + ... + \alpha_{m-1}\tilde{p}_{i+1} + \alpha_m \tilde{p}_i = 0.$$

  Because $\alpha_m \tilde{p}_i \neq 0$, there exists $j \in \{i+1, ..., i+m\}$ such that $\tilde{p}_j \neq 0$.

- If $\alpha_m = 0$, then $\alpha_{m-1} \neq 0$, and we know again from equation (4) that

  $$\tilde{p}_{i+m-1} + \alpha_1 \tilde{p}_{i+m-2} + ... + \alpha_{m-1}\tilde{p}_i = 0.$$

  Because $\alpha_{m-1}\tilde{p}_i \neq 0$, there exists $j \in \{i+1, ..., i+m-1\}$ such that $\tilde{p}_j \neq 0$.

So we know that there exists $j > i$ such that $\tilde{p}_j = \mathbb{E}(p_j) \neq 0$. This directly implies $p_j \neq 0$. Now let us assume that $\{k \in \mathbb{N} \mid p_k \neq 0\}$ is finite. Hence a maximum m of this set exists. As we showed above, there exists $j > m$ such that $p_j \neq 0$, which implies $j \in \{k \in \mathbb{N} \mid p_k \neq 0\}$. But then the maximum of this set is smaller than j, which is a contradiction. Thus, the lemma applies. □

Now we transfer the notion pseudo-regularity to endomorphisms.

Definition 3.9. Let M be a free finite-dimensional R-module with basis $B = (b_1, ..., b_n)$ and $\psi : M \to M$ be an endomorphism. Then ψ is called *pseudo-regular* if the transformation matrix $M_{B,B}(\psi)$ is pseudo-regular.

This definition is independent of the chosen basis B. Indeed let B and B' be two bases of M. If $M_{B,B}(\psi)$ is pseudo-regular, we know that the minimal polynomial $P(\lambda)$ of the matrix $\mathbb{E}(M_{B,B}(\psi))$ does not divide λ^2. Since $M_{B',B'}(\psi)$ is similar to $M_{B,B}(\psi)$, we also know that $\mathbb{E}(M_{B',B'}(\psi))$ is similar to $\mathbb{E}(M_{B,B}(\psi))$ and has the same minimal polynomial (Lemma 2.17), which satisfies the required pseudo-regularity condition.

Now we are able to state the above lemma again, but this time over endomorphisms instead of matrices.

Lemma 3.11. *Let M be a free finite-dimensional R-module and ϕ an endomorphism on M. Further, let $\lambda : M \to R$ be a linear form and $\gamma \in M$. We define the sequence $(p_k)_{k\in\mathbb{N}}$ by $p_k := (\lambda \circ \phi^{(k)})(\gamma)$. If ϕ is pseudo-regular and $p_1 \neq 0$, then the set $\{k \in \mathbb{N} \mid p_k \neq 0\}$ is infinite.*

Proof. Let $B = (b_1, ..., b_n)$ be an arbitrary basis of M. We take (1) as basis of the free R-module R. Assume that ϕ is pseudo-regular and $p_1 \neq 0$. Then $M_{B,B}(\phi)$ is pseudo-regular and

$$M_{B,(1)}(\lambda) \cdot M_{B,B}(\phi) \cdot \Omega_B(\gamma) = (\lambda \circ \phi)(\gamma) = p_1 \neq 0.$$

Hence, we know by Lemma 3.10 that

$$M_{B,(1)}(\lambda) \cdot M_{B,B}(\phi)^k \cdot \Omega_B(\gamma) \neq 0$$

for infinitely many k since $M_{B,(1)} \in R^{1\times n}, M_{B,B} \in R^{n\times n}$ and $\Omega_B(\gamma) \in R^{n\times 1}$. This directly implies

$$(\lambda \circ \phi^{(k)})(\gamma) \neq 0$$

for infinitely many k. □

The generalization of the previous theorem is done in a straightforward way by using the multiplicativity of the embedding function over matrices.

Theorem 5. *Let $n \in \mathbb{N}$. There exists $N \in \mathbb{N}$ such that for all $(A_i)_{1\leq i\leq N} \in (R^{n\times n})^{\{1,...,N\}}$ there exist $1 \leq i \leq j \leq N$ such that $A_i...A_j$ is a pseudo-regular matrix.*

Proof. We already know by Theorem 4 that there exist $1 \leq i \leq j \leq N$ such that $\mathbb{E}(A_i)...\mathbb{E}(A_j)$ is a pseudo-regular matrix (with respect to Definition 3.1). $\mathbb{E}\big|_{R^{n\times n}}$ is a multiplicative function. Hence

$$\mathbb{E}(A_i)...\mathbb{E}(A_j) = \mathbb{E}(A_i...A_j),$$

which implies that $\mathbb{E}(A_i...A_j)$ is pseudo-regular (by Definition 3.1). Therefore we obtain the pseudo-regularity of $A_i...A_j$ (with respect to the definition over integral domains). □

Note that our approach of generalization does not work for arbitrary commutative rings. We need commutative rings in which the cancellation law holds, that is, integral domains, in order to embed the commutative ring in a field. That means we can only embed integral domains in a field, namely, the field of fractions.

4 Formal power series on trees

In this chapter, we consider fundamental definitions for *formal power series* on trees. We also determine when such a formal power series is called *recognizable*. We generalize the definitions from Berstel and Reutenauer [BR80] to integral domains and modules over them. Therefore, let R be an integral domain in the following.

We use the notation

$$\mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_1 \cup ...$$

for the set of function symbols. The sets $\mathcal{F}_i, i \geq 1$, are disjoint. Therefore, we are able to define the **arity** of a function symbol $f \in \mathcal{F}$ by

$$\text{arity}(f) = p :\Leftrightarrow f \in \mathcal{F}_p.$$

Moreover, we assume that $\mathcal{F}$ is a finite set in the following.

Now we need to clarify what *trees* are. We define them over the set of function symbols $\mathcal{F}$ inductively.

Definition 4.1 (tree)**.** Let $\mathcal{T}(\mathcal{F})$ be the smallest set with

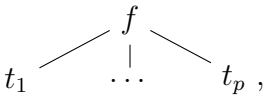
- $f_0 \in \mathcal{T}(\mathcal{F})$ for arbitrary $f_0 \in \mathcal{F}_0$ and
- $p \geq 1, f \in \mathcal{F}_p, t_1, ..., t_p \in \mathcal{T}(\mathcal{F}) \Rightarrow f(t_1, ..., t_p) \in \mathcal{T}(\mathcal{F})$.

$\mathcal{T}(\mathcal{F})$ is called the *set of trees*. Every $t \in \mathcal{T}(\mathcal{F})$ is called a *tree*.

We can represent a tree by the term

$$f(t_1, ..., t_p),$$

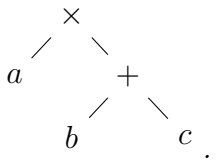
whereby $t_1, ..., t_p$ are represented in the same way, recursively. Another notation for this is



whereby $t_1, ..., t_p$ are also recursively notated like this.

Trees can be used to represent arithmetic expressions for instance. Berstel and Reutenauer discussed this example in detail [BR80]. We consider the example step by step as we move forward with our definitions and theorems.

Example. *We define* $\mathcal{F}_0 := \{a, b, c, ..., z\}, \mathcal{F}_1 := \{-\}, \mathcal{F}_2 := \{+, \times\}$ *and* $\mathcal{F} := \mathcal{F}_0 \cup \mathcal{F}_1 \cup \mathcal{F}_2$*. Then we can represent arithmetic expressions like* $a \times (b + c)$ *as*



We will come back to this example after we introduce formal power series on trees.

Definition 4.2 (formal power series)**.** A *formal power series* on $\mathcal{T}(\mathcal{F})$ is a function

$$S : \mathcal{T}(\mathcal{F}) \to R.$$

The *set of all formal power series* on $\mathcal{T}(\mathcal{F})$ is denoted by $R\langle\langle\mathcal{F}\rangle\rangle$. We call $S(t)$ the *coefficient* of t in S. Further, S can be notated as

$$S = \sum_{t \in \mathcal{T}(\mathcal{F})} S(t)t.$$

Now let us come back to our previously stated example. We defined trees that represent arithmetic expressions. Now we want to evaluate these expressions by assigning each arithmetic expression a value. This can be done by a formal power series on these trees.

Example. *$\mathcal{F}$ is defined as before. Now we define a formal power series on $\mathcal{T}(\mathcal{F})$ in the following way: Let*

$$f : \mathcal{F}_0 \to \mathbb{Q}$$

be an arbitrary mapping that assigns each variable in $\mathcal{F}_0$ a rational number (which values are concretely assigned does not matter in the following). Then we extend this mapping in a natural way. Indeed we apply the functions occurring in the tree to these assigned numbers. This is done by the function $S : \mathcal{T}(\mathcal{F}) \to \mathbb{Q}$ defined as

$$S(t) := \begin{cases} f(t) & \text{if } t \in \mathcal{F}_0 \\ -S(t) & \text{if } t = -(t_1) \text{ for some } t_1 \in \mathcal{T}(\mathcal{F}) \\ S(t_1) + S(t_2) & \text{if } t = +(t_1, t_2) \text{ for some } t_1, t_2 \in \mathcal{T}(\mathcal{F}) \\ S(t_1) \cdot S(t_2) & \text{if } t = \times(t_1, t_2) \text{ for some } t_1, t_2 \in \mathcal{T}(\mathcal{F}). \end{cases}$$

That is, every tree is evaluated recursively by this function and is assigned a rational number, namely, the solution of the arithmetic expression, whereby the variables are exchanged by their values. For instance, we obtain with $f(a) := 1, f(b) := 2, f(c) := 3$ the recursive evaluation

$$\begin{aligned} S\left(\begin{array}{ccccc} & & \times & & \\ & \diagup & & \diagdown & \\ a & & & & + \\ & & & \diagup & \diagdown \\ & & b & & c \end{array}\right) &= S(a) \cdot S\left(\begin{array}{ccc} & + & \\ \diagup & & \diagdown \\ b & & c \end{array}\right) \\ &= S(a) \cdot (S(b) + S(c)) \\ &= f(a) \cdot (f(b) + f(c)) \\ &= 1 \cdot (2 + 3) \\ &= 5. \end{aligned}$$

Again we continue with a few definitions and will come back to this example later.

A formal power series can also generate a language on trees, the so-called *support*.

Definition 4.3 (support)**.** Let S be a formal power series on $\mathcal{T}(\mathcal{F})$. Then the *support* of S is defined by

$$\text{supp}(S) := \{t \in \mathcal{T}(\mathcal{F}) \mid S(t) \neq 0\}.$$

If $\text{supp}(S)$ is finite, we call S a *polynomial*. $R\langle\mathcal{F}\rangle$ denotes the *set of all polynomials*.

$(R\langle\langle\mathcal{F}\rangle\rangle, +, \cdot)$, whereby $+$ and $\cdot$ are addition and scalar multiplication on functions, is a module. It is easy to show that $R\langle\mathcal{F}\rangle \subseteq R\langle\langle\mathcal{F}\rangle\rangle$ is a submodule of $R\langle\langle\mathcal{F}\rangle\rangle$.

Next we define the notion *recognizable* for formal power series. This is done by mapping trees to vectors.

Let M be a free finite-dimensional R-module in the following.

$$\mathcal{L}(M^p, M)$$

denotes the set of all multilinear functions from M^p to M for $p \geq 1$. Then we define

$$\mathcal{L} := \bigcup_{p \geq 0} \mathcal{L}(M^p, M),$$

whereby we set $\mathcal{L}(M^0, M) := M$.

Each function symbol can now be represented as such a multilinear function.

Definition 4.4 (linear representation)**.** (M, μ) is called a *linear representation* of $\mathcal{T}(\mathcal{F})$ if M is a free finite-dimensional R-module and μ is a function

$$\mu : \mathcal{F} \to \mathcal{L}$$

with

$$\mu(f) = l \in \mathcal{L}(M^p, M) \Leftrightarrow f \in \mathcal{F}_p.$$

We denote $\bar{\mu}$ as the natural extension of μ to $\mathcal{T}(\mathcal{F})$ that is a function

$$\bar{\mu} : \mathcal{T}(\mathcal{F}) \to M$$

with

$$\bar{\mu}(t) := \begin{cases} \mu(t) & \text{if } t \in \mathcal{F}_0 \\ \mu(f)(\bar{\mu}(t_1), ..., \bar{\mu}(t_p)) & \text{if } t = f(t_1, ..., t_p), f \in \mathcal{F}_p, t_1, ..., t_p \in \mathcal{T}(\mathcal{F}), p \geq 1. \end{cases}$$

$\bar{\mu}$ is unique and well-defined because the representation $t = f(t_1, ..., t_p)$ is unique since $\mathcal{T}(\mathcal{F})$ is freely generated by $\mathcal{F}$. Moreover, we know that

$$\bar{\mu}\big|_{\mathcal{F}_0} = \mu\big|_{\mathcal{F}_0}.$$

Hence, we no longer distinguish between μ and $\bar{\mu}$, and will use the same symbol μ for both functions from now on.

Definition 4.5 (recognizable formal power series)**.** Let S be a formal power series on $\mathcal{T}(\mathcal{F})$. Then S is called *recognizable* if there exists (M, μ, λ), whereby (M, μ) is a linear representation of $\mathcal{T}(\mathcal{F})$ and

$$\lambda : M \to R$$

is a linear function with

$$S(t) = \lambda(\mu(t))$$

for every $t \in \mathcal{T}(\mathcal{F})$.

This means in words that we are able to calculate function values of S by applying multilinear functions over a module. We will see that this is helpful in the proof of the iteration theorem since it helps us to use transformation matrices and to apply statements about pseudo-regular matrices.

Fortunately, our previously discussed example about arithmetic expressions is a recognizable formal power series.

Example. *We define $\mathcal{F}$ and S as before. Then we need to find such a tuple (M, μ, λ) as in the definition above, whereby (M, μ) is a linear representation of $\mathcal{T}(\mathcal{F})$. We choose M to be $\mathbb{Q}^2$, which is a module, and even a vector space, over the rational numbers. Now we need to define the function μ that maps our function symbols to multilinear functions. Indeed we define it in the following way:*

$$\begin{aligned}
\mu(a) &:= e_1 + S(a)e_2 \text{ if } a \in \mathcal{F}_0,\\
\mu(-)(e_1) &:= e_1, \mu(-)(e_2) := -e_2,\\
\mu(+)(e_1, e_1) &:= e_1, \mu(+)(e_1, e_2) = \mu(+)(e_2, e_1) := e_2, \mu(+)(e_2, e_2) := 0,\\
\mu(\times)(e_1, e_1) &:= e_1, \mu(\times)(e_1, e_2) = \mu(\times)(e_2, e_1) := 0 \text{ and } \mu(\times)(e_2, e_2) := e_2,
\end{aligned}$$

whereby (e_1, e_2) is the canonial basis of $\mathbb{Q}^2$. At last we define the linear form λ by

$$\lambda(e_1) := 0 \text{ and } \lambda(e_2) = 1.$$

Note that it is enough to determine the image of all combinations of basis vectors to uniquely define a multilinear mapping on vector spaces. The idea behind these definitions is to keep the first entry in the vector always 1 and to do the calculation in the second entry. Now, somebody could wonder why we actually need the first entry and why it is not enough to use one-dimensional vectors, that is, using $\mathbb{Q}$ as module. The reason for this is that μ needs to assign multilinear functions. Indeed if we would define $\mu(+)$ as the common addition for instance, this function would not be multilinear. Therefore, we need two dimensions to obtain multilinearity. Now let us verify that $\mu(-)$, $\bar{+} := \mu(+)$ and $\bar{\times} := \mu(\times)$ actually operate like the common inversion, addition and multiplication. We have

$$\mu(a) := \begin{pmatrix} 1 \\ S(a) \end{pmatrix} \text{ if } a \in \mathcal{F}_0,$$

$$\mu(-)\begin{pmatrix} 1 \\ x \end{pmatrix} = \mu(-)(e_1 + xe_2) = \mu(-)(e_1) + x\mu(-)(e_2) = e_1 - xe_2 = \begin{pmatrix} 1 \\ -x \end{pmatrix},$$

$$\begin{aligned}
\begin{pmatrix}1\\x\end{pmatrix} \bar{+} \begin{pmatrix}1\\y\end{pmatrix} &= (e_1 + xe_2)\bar{+}(e_1 + ye_2) \\
&= (e_1\bar{+}e_1) + (e_1\bar{+}ye_2) + (xe_2\bar{+}e_1) + (xe_2\bar{+}ye_2) \\
&= e_1 + y(e_1\bar{+}e_2) + x(e_2\bar{+}e_1) + xy(e_2\bar{+}e_2) \\
&= e_1 + ye_2 + xe_2 + 0 \\
&= e_1 + (x+y)e_2 \\
&= \begin{pmatrix}1\\x+y\end{pmatrix}
\end{aligned}$$

and at last

$$\begin{aligned}
\begin{pmatrix}1\\x\end{pmatrix} \bar{\times} \begin{pmatrix}1\\y\end{pmatrix} &= (e_1 + xe_2)\bar{\times}(e_1 + ye_2) \\
&= (e_1\bar{\times}e_1) + (e_1\bar{\times}ye_2) + (xe_2\bar{\times}e_1) + (xe_2\bar{\times}ye_2) \\
&= (e_1\bar{\times}e_1) + y(e_1\bar{\times}e_2) + x(e_2\bar{\times}e_1) + xy(e_2\bar{\times}e_2) \\
&= e_1 + 0 + 0 + xye_2 \\
&= e_1 + xye_2 \\
&= \begin{pmatrix}1\\xy\end{pmatrix}
\end{aligned}$$

for all $x, y \in \mathbb{Q}$. *Hence we know that*

$$\lambda(\mu(t)) = \lambda\left(\begin{pmatrix}1\\S(t)\end{pmatrix}\right) = \lambda(e_1 + S(t)e_2) = \lambda(e_1) + S(t)\lambda(e_2) = 0 + S(t) = S(t),$$

which means that the formal power series S *is recognizable.*

We are still not done with this example yet. We again refer to it in the next chapter.

5 The generalized iteration theorem

Berstel and Reutenauer [BR80] proved the iteration theorem for recognizable formal power series on trees over fields and vector spaces. It states that we can repeat a part of a long enough walk in a tree taken from the support of a recognizable formal power series such that the tree is still a member of the support for infinitely many such iterations. But this does not only hold over fields and vector spaces, it still holds over integral domains and modules over integral domains. Indeed we use Berstel's and Reuternauer's approach and generalize it in this chapter. R denotes an integral domain in the following.

We first need to describe a walk in a tree to be able to state the iteration theorem. As in the previous chapter, $\mathcal{F}$ denotes the set of function symbols and $\mathcal{F}_p$ the set of function symbols with arity p. Now we define $\mathcal{F}' := \mathcal{F} \cup \{x\}$, whereby x is a new symbol with arity 0. That is, $x \notin \mathcal{F}$. We also define $\mathcal{F}'_0 := \mathcal{F}_0 \cup \{x\}$. Next we need two functions. The first is defined for arbitrary $\tilde{t} \in \mathcal{T}(\mathcal{F}')$ by

$$\psi_{\tilde{t}} : \mathcal{T}(\mathcal{F}') \to \mathcal{T}(\mathcal{F}')$$

with

$$\psi_{\tilde{t}}(t) := \begin{cases} \tilde{t} & \text{if } t = x \\ f_0 & \text{if } t = f_0 \in \mathcal{F}_0 \\ f(\psi_{\tilde{t}}(t_1), ..., \psi_{\tilde{t}}(t_p)) & \text{if } t = f(t_1, ..., t_p), p \geq 1, f \in \mathcal{F}_p, t_1, ..., t_p \in \mathcal{T}(\mathcal{F}'). \end{cases}$$

The definition tells us that $\psi_{\tilde{t}}$ takes a tree over $\mathcal{T}(\mathcal{F}')$ as input and replaces every occurrence of x with $\tilde{t}$. The second function

$$\bar{\phi}_s : \mathcal{T}(\mathcal{F}') \to \mathcal{T}(\mathcal{F}')$$

is defined for arbitrary $s \in \mathcal{T}(\mathcal{F}')$ by

$$\bar{\phi}_s(t) := \psi_t(s).$$

Hence, $\bar{\phi}_s$ takes a tree in $\mathcal{T}(\mathcal{F}')$ as input and replaces every occurrence of x in s with this tree. Because we only want to allow replacing x with a tree with no occurrence of x, we restrict $\bar{\phi}_s$ to $\mathcal{T}(\mathcal{F})$. Then $\bar{\phi}_s$ only maps to trees that do not contain the symbol x since we replace each occurrence with a tree taken from $\mathcal{T}(\mathcal{F})$. Therefore, we define

$$\phi_s : \mathcal{T}(\mathcal{F}) \to \mathcal{T}(\mathcal{F}), \phi_s(t) := \bar{\phi}_s(t).$$

Next, let A and B be two sets defined by

$$A := \{s \in \mathcal{T}(\mathcal{F}') \mid \#_x s = 1\}$$

and

$$B := \{s \in \mathcal{T}(\mathcal{F}') \mid s = f(t_1, ..., t_p), f \in \mathcal{F}_p, \\ \exists! 1 \leq i \leq p : t_i = x, t_j \in \mathcal{T}(\mathcal{F}) \text{ for } j \neq i\}.$$

Now the idea is to construct every tree taken from A with trees taken from B. Then this step by step construction describes a walk in the tree taken from A. We need to formalize this by using our replacing functions. We define

$$\Sigma^* := \{\phi_s \mid s \in A\} \text{ and } \Sigma := \{\phi_s \mid s \in B\}.$$

As the notion indicates, Σ^* is freely generated by Σ. **Freely** means that every element in Σ^* has a unique decomposition with elements taken from Σ up to concatenation with the identity element.

Proposition 5.1. *Σ^* is a monoid freely generated by Σ. The neutral element is ϕ_x.*

Proof. At first we show that Σ is a generator of Σ^*. We do this by induction on the height of the trees. We define the height of a tree recursively by

$$\text{height} : \mathcal{T}(\mathcal{F}') \to \mathbb{N}, \text{height}(f(t_1, ..., t_p)) := \max_{i=1,...,p}(\text{height}(t_i)) + 1,$$

whereby we set $\text{height}(t) := 1$ for all $t \in \mathcal{F}'_0$. Note that ϕ_x is generated by the empty composition. Therefore we start our induction with $\phi_s \in \Sigma^*$, whereby $\text{height}(s) = 2$. $\#_x s = 1$ and since x is a symbol with arity 0, we know that

$$s = f(t_1, ..., t_{i-1}, x, t_{i+1}, ..., t_p)$$

with $t_1, ..., t_{i-1}, t_{i+1}, ..., t_p \in \mathcal{F}_0$ (which means unequal x). Therefore $\phi_s \in \Sigma$ and we are done. Now let $s \in A \setminus (B \cup \{x\})$ with $\text{height}(s) = j + 1$ and assume that every $\phi_{\tilde{s}}$ with $\text{height}(\tilde{s}) \leq j$ decomposes as desired. Then we have

$$s = f(t_1, ..., t_p),$$

whereby $f \in \mathcal{F}_p$ and it exists exactly one $1 \leq i \leq p$ such that $t_i \in A$ and $t_j \in \mathcal{T}(\mathcal{F})$ for all $j \neq i$. This means that exactly one child-tree of the node f has an occurrence of x. As a consequence, we are able to state

$$\bar{s} := f(t_1, ..., t_{i-1}, x, t_{i+1}, ..., t_p) \in B,$$

hence

$$\phi_{\bar{s}} \in \Sigma.$$

We obtain the composition

$$\phi_s = \phi_{\bar{s}} \circ \phi_{t_i},$$

which indeed is a result of

$$\begin{aligned}\phi_s(t) &= \psi_t(s) = \psi_t(f(t_1, ..., t_p)) = f(t_1, ..., t_{i-1}, \psi_t(t_i), t_{i+1}, ..., t_p) \\ &= f(t_1, ..., t_{i-1}, \phi_{t_i}(t), t_{i+1}, ..., t_p) = \psi_{\phi_{t_i}(t)}(f(t_1, ..., t_{i-1}, x, t_{i+1}, ..., t_p)) \\ &= \psi_{\phi_{t_i}(t)}(\bar{s}) = \phi_{\bar{s}}(\phi_{t_i}(t)) = (\phi_{\bar{s}} \circ \phi_{t_i})(t)\end{aligned}$$

for arbitrary $t \in \mathcal{T}(\mathcal{F})$. It remains to be proven that $\text{height}(t_i) \leq j$ to conclude by induction. We verify this by

$$\text{height}(t_i) \leq \max_{k=1,...,p} \text{height}(t_k) = \text{height}(f(t_1, ..., t_p)) - 1 = (j + 1) - 1 = j.$$

Next we need to show that Σ generates freely, that is, every decomposition is unique up to composition with ϕ_x. Therefore assume that $\phi_s \in \Sigma^*$ has two decompositions

$$\phi_{\sigma_1} \circ ... \circ \phi_{\sigma_n} = \phi_s = \phi_{\tau_1} \circ ... \circ \phi_{\tau_m}.$$

We proceed by induction on n. If $n = 0$, we know that $\phi_s = \phi_x$, which decomposes uniquely as the empty composition, and $n = m$. Now let $n \in \mathbb{N}_+$ and assume that every decomposition of length $n-1$ is unique. Then let $t \in \mathcal{T}(\mathcal{F})$ with $\text{height}(t) > \text{height}(\sigma_1), ..., \text{height}(\sigma_n), \text{height}(\tau_1), ..., \text{height}(\tau_m)$. In particular we have

$$(\phi_{\sigma_1} \circ ... \circ \phi_{\sigma_n})(t) = (\phi_{\tau_1} \circ ... \circ \phi_{\tau_m})(t).$$

Because $\sigma_1, \tau_1 \in B$, we obtain

$$\sigma_1 = f(s_1, ..., s_{i-1}, x, s_{i+1}, ..., s_p)$$

and

$$\tau_1 = g(t_1, ..., t_{j-1}, x, t_{j+1}, ..., t_q)$$

for some $1 \leq i \leq p, 1 \leq j \leq q, s_1, ..., s_{i-1}, s_{i+1}, ..., s_p, t_1, ..., t_{j-1}, t_{j+1}, ..., t_q \in \mathcal{T}(\mathcal{F})$. Further, we define

$$u := \phi_{\sigma_2} \circ ... \circ \phi_{\sigma_n}(t)$$

and

$$v := \phi_{\tau_2} \circ ... \circ \phi_{\tau_m}(t).$$

Then

$$\phi_{\sigma_1}(u) = \phi_{\tau_1}(v),$$

hence

$$f(s_1, ..., s_{i-1}, u, s_{i+1}, ..., s_p) = g(t_1, ..., t_{j-1}, v, s_{j+1}, ..., t_q).$$

It is trivial that it has to be $f = g$ and

$$(s_1, ..., s_{i-1}, u, s_{i+1}, ..., s_p) = (t_1, ..., t_{j-1}, v, s_{j+1}, ..., t_q).$$

We know that u and v have the same position. The position cannot differ because for all $k \neq j$

$$\text{height}(u) \geq \text{height}(t) > \text{height}(\tau_1) > \text{height}(t_k) \Rightarrow u \neq t_k.$$

So $i = j$ and $s_k = t_k$ for all $k \neq i$, which means $\sigma_1 = \tau_1$. Moreover $u = v$, that is,

$$\phi_{\sigma_2} \circ ... \circ \phi_{\sigma_n}(t) = \phi_{\tau_2} \circ ... \circ \phi_{\tau_m}(t).$$

Hence, we conclude by induction. □

This proposition helps us to define a walk in a tree since we know now that we are able to construct each tree in A step by step with trees taken from B.

Definition 5.1 (walk)**.** Let $t \in \mathcal{T}(\mathcal{F})$. A *walk* in t is a pair (ϕ, a) with $\phi \in \Sigma^*$ and $a \in \mathcal{F}_0$ such that $t = \phi(a)$. The *length* of this walk is the length of ϕ in the free monoid Σ^*.

So a walk describes a path from the root node to a leaf node. Also note that the length of a walk is well-defined since the decomposition of ϕ into elements of Σ is unique.

We denote by $f^{(k)}$ the k-fold composition of f in the following, that is, $f \circ ... \circ f$.

Theorem 6 (Iteration Theorem for Recognizable Formal Power Series on Trees)**.** *Let $S \in R\langle\langle \mathcal{F} \rangle\rangle$ be recognizable. Then there exists $N \in \mathbb{N}$ such that for an arbitrary $t \in \operatorname{supp}(S)$ and every walk (ϕ, a) in t with length greater or equal to N, there exists a decomposition of ϕ into $\phi_1 \circ \phi_2 \circ \phi_3$ in Σ^* such that*

$$\{(\phi_1 \circ \phi_2^{(k)} \circ \phi_3)(a) \mid k \in \mathbb{N}\} \cap \operatorname{supp}(S)$$

is an infinite set.

Proof. Let (M, μ, λ) be a tuple such that (M, μ) is a linear representation of $\mathcal{T}(\mathcal{F})$ and $\lambda : M \to R$ is a linear form that satisfies $\lambda(\mu(t)) = S(t)$ for all $t \in \mathcal{T}(\mathcal{F})$. We first want to translate the functions ϕ_t from above to modules. We do this by defining $\hat{\phi}_\sigma : M \to M$ for $\sigma = f(s_1, ..., s_{i-1}, x, s_{i+1}, ..., s_p) \in B$ as

$$\hat{\phi}_\sigma(v) := \mu(f)(\mu(s_1), ..., \mu(s_{i-1}), v, \mu(s_{i+1}), ..., \mu(s_p)).$$

We extend the definition to $\phi_\tau \in \Sigma^*$ with $\tau \in A$ by composition. This means that we take the unique decomposition of ϕ_τ in Σ^*, $\phi_\tau = \phi_{\sigma_1} \circ ... \circ \phi_{\sigma_q}$, and define $\hat{\phi}_\tau : M \to M$ as

$$\hat{\phi}_\tau := \hat{\phi}_{\sigma_1} \circ ... \circ \hat{\phi}_{\sigma_q}.$$

This leads to the desired translation in the form of a mapping

$$\phi \mapsto \hat{\phi}$$

for all $\phi \in \Sigma^*$ that satisfies

$$\hat{\phi}(\mu(t)) = \mu(\phi(t))$$

for all $t \in \mathcal{T}(\mathcal{F})$. The equation is verified by induction on the length of the decomposition. The assertion is clear for decompositions of length 0, that is, $\phi = \phi_x$ since $\hat{\phi}_x = \mathrm{Id}_M$ and $\phi_x(t) = t$. We want to start our induction with decompositions of length 1, that means $\phi_\sigma \in \Sigma$ with $\sigma = f(s_1, ..., s_{i-1}, x, s_{i+1}, ..., s_p) \in B$. Then we have

$$\begin{aligned}\hat{\phi}_\sigma(\mu(t)) &= \mu(f)(\mu(s_1), ..., \mu(s_{i-1}), \mu(t), \mu(s_{i+1}), ..., \mu(s_p)) \\ &= \mu(f(s_1, ..., s_{i-1}, t, s_{i+1}, ..., s_p)) \\ &= \mu(\phi_\sigma(t)).\end{aligned}$$

If the equation holds for decompositions of length j $(*)$, we obtain with

$$\phi_\tau = \phi_{\sigma_1} \circ ... \circ \phi_{\sigma_{j+1}} \in \Sigma^*$$

for arbitrary $\phi_{\sigma_i} \in \Sigma, i \in \{1, ..., j+1\}$, and $\sigma_1 = g(t_1, ..., t_{i-1}, x, t_{i+1}, ..., t_q)$ that

$$\begin{aligned}
\hat{\phi}_\tau(\mu(t)) &= (\hat{\phi}_{\sigma_1} \circ ... \circ \hat{\phi}_{\sigma_{j+1}})(\mu(t)) \\
&= \hat{\phi}_{\sigma_1}((\underbrace{\hat{\phi}_{\sigma_2} \circ ... \circ \hat{\phi}_{\sigma_{j+1}}}_{\phi := \phi_{\sigma_2} \circ ... \circ \phi_{\sigma_{j+1}}})(\mu(t))) \\
&= \hat{\phi}_{\sigma_1}(\hat{\phi}(\mu(t))) \\
&\overset{(*)}{=} \hat{\phi}_{\sigma_1}(\mu(\phi(t))) \\
&= \mu(g)(\mu(t_1), ..., \mu(t_{i-1}), \mu(\phi(t)), \mu(t_{i+1}), ..., \mu(t_q)) \\
&= \mu(g(t_1, ..., t_{i-1}, \phi(t), t_{i+1}, ..., t_q)) \\
&= \mu(\phi_{\sigma_1}(\phi(t))) \\
&= \mu((\phi_{\sigma_1} \circ ... \circ \phi_{\sigma_{j+1}})(t)) \\
&= \mu(\phi_\tau(t)).
\end{aligned}$$

Next, we choose N in dependency of $n := \dim(M)$ as in Theorem 5. Let $t \in \text{supp}(S)$ and (ϕ, a) be an arbitrary walk in t with length greater or equal to N. Therefore we know that there exist unique $\phi_{\sigma_1}, ..., \phi_{\sigma_k} \in \Sigma$, $k \geq N$ such that

$$\phi = \phi_{\sigma_1} \circ ... \circ \phi_{\sigma_k},$$

hence

$$\hat{\phi} = \hat{\phi}_{\sigma_1} \circ ... \circ \hat{\phi}_{\sigma_k}.$$

We take an arbitrary basis $B := (b_1, ..., b_n)$ of M and denote by A_i the transformation matrix of the endomorphism $\hat{\phi}_{\sigma_i}$ with basis B for all $1 \leq i \leq k$. Then we obtain a family

$$(A_1, ..., A_k)$$

of square matrices, which are members of $R^{n \times n}$. We know by applying Theorem 5 that there exist $1 \leq i \leq j \leq N$ such that $A_i ... A_j$ is pseudo-regular. The transformation matrix A of ϕ with basis B satisfies

$$A = A_1 \cdot ... \cdot A_k = A_1 \cdot ... \cdot A_{i-1} \cdot A_i \cdot ... \cdot A_j \cdot A_{j+1} \cdot ... \cdot A_k.$$

We define

$$\begin{aligned}
\phi_1 &= \phi_{\sigma_1} \circ ... \circ \phi_{\sigma_{i-1}}, \\
\phi_2 &= \phi_{\sigma_i} \circ ... \circ \phi_{\sigma_j}, \\
\phi_3 &= \phi_{\sigma_{j+1}} \circ ... \circ \phi_{\sigma_k}.
\end{aligned}$$

Then we obtain the decomposition

$$\phi = \phi_1 \circ \phi_2 \circ \phi_3,$$

whereby $\hat{\phi}_2$ is a pseudo-regular endomorphism. We set

$$p_m := (\lambda \circ \hat{\phi}_1 \circ \hat{\phi}_2^{(m)} \circ \hat{\phi}_3)(\mu(a))$$

for $m \geq 0$. It is the case that

$$p_1 = (\lambda \circ \hat{\phi})(\mu(a)) = \lambda(\hat{\phi}(\mu(a))) = \lambda(\mu(\phi(a))) = \lambda(\mu(t)) = S(t) \neq 0,$$

because t is a member of the support. Moreover, $a \in \mathcal{F}_0 \Rightarrow \mu(a) \in M \Rightarrow \hat{\phi}_3(\mu(a)) \in M$, $\lambda \circ \hat{\phi}_1 : M \to R$ is a linear form and $\hat{\phi}_2^{(m)}$ is a pseudo-regular endomorphism on M. Hence, we can apply Lemma 3.11 and know for infinitely many $m \geq 0$ that $p_m \neq 0$. This means for these m that

$$\begin{aligned} p_m &= (\lambda \circ \hat{\phi}_1 \circ \hat{\phi}_2^{(m)} \circ \hat{\phi}_3)(\mu(a)) \\ &= \lambda((\hat{\phi}_1 \circ \hat{\phi}_2^{(m)} \circ \hat{\phi}_3)(\mu(a))) \\ &= \lambda(\mu((\phi_1 \circ \phi_2^{(m)} \circ \phi_3)(a))) \\ &= (\lambda \circ \mu)((\phi_1 \circ \phi_2^{(m)} \circ \phi_3)(a)) \\ &= S((\phi_1 \circ \phi_2^{(m)} \circ \phi_3)(a)) \\ &\neq 0. \end{aligned}$$

Therefore, we know for infinitely many m that

$$(\phi_1 \circ \phi_2^{(m)} \circ \phi_3)(a) \in \operatorname{supp}(S),$$

hence

$$\{\phi_1 \circ \phi_2^{(k)} \circ \phi_3 \mid k \in \mathbb{N}\} \cap \operatorname{supp}(S)$$

is an infinite set. □

Let us again come back to our example concerning arithmetic expressions. We already know that this formal power series is recognizable. Indeed, there is a practical implication of the above theorem. It states that if a long enough arithmetic expression (without division) is unequal to zero, which means it is contained in the support, then we are able to repeat some part of the arithmetic expression such that it is still unequal to zero.

Example. *We know that $a+(-(-(-b))) \neq 0$ for all rational numbers $a \neq b$. Thus, we assume $S(a) \neq S(b)$ for this example. Then we have*

$$S\underbrace{\left(\begin{array}{ccc} & + & \\ a & & - \\ & & | \\ & & - \\ & & | \\ & & - \\ & & | \\ & & b \end{array}\right)}_{=:t} \neq 0.$$

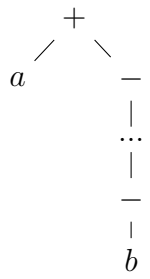
We know that t has a walk (ϕ, b), whereby ϕ decomposes in $\phi = \phi_{t_1} \circ \phi_{t_2} \circ \phi_{t_2} \circ \phi_{t_2}$ with

$$t_1 := \begin{array}{ccc} & + & \\ a & & x \end{array} \quad \text{and} \quad t_2 := \begin{array}{c} - \\ | \\ x \end{array},$$

because $\phi(b) = t$. Since the length of this walk is 4, which is greater than 3, which is the required length of the above theorem, we know that there exists a part of the decomposition, which we can repeat such that the tree is still a member of the support, hence the arithmetic expression is still unequal to zero. Indeed this is actually true. Since

$$a + \overbrace{(-(\ldots -}^{2k+1 \text{ times}}(b))) = a + (-b) \neq 0$$

for all $k \in \mathbb{N}$, we know that S still does not map the tree



to zero, whereby $-$ occurs odd number of times in the tree. Hence

$$\phi_{t_1} \circ \phi_{t_2}^{(2k+1)} \circ \phi_{t_2} \circ \phi_{t_2} \in \operatorname{supp}(S)$$

for arbitrary $k \in \mathbb{N}$.

This example also shows us that this repeatable part is not unique. Indeed we can also state that

$$\phi_{t_1} \circ (\phi_{t_2} \circ \phi_{t_2})^{(k)} \circ \phi_{t_2} \in \operatorname{supp}(S)$$

for all $k \in \mathbb{N}$.

6 Conclusion

The aim of this thesis was to generalize Berstel's and Reutenauer's iteration theorem for recognizable formal power series on trees. We proved that it holds over arbitrary integral domains and modules over them. The main idea of this proof was to construct the field of fractions of the integral domain and apply the known statements over fields.

6.1 Future work

It is still unclear whether the iteration theorem holds over arbitrary commutative rings. The approach to construct the field of fractions does not work in this case since we used the cancellation law in the proof of Lemma 2.8. Therefore, we need a way to define pseudo-regularity of matrices over a commutative ring. The problem with this is that many of the equivalent conditions in Proposition 3.1 do not exist. The reason for this is that the rank and also the minimal polynomial of a matrix with entries taken from a commutative ring are no longer well-defined. These conditions are frequently used throughout this thesis. Therefore, the approach to prove that every long enough matrix-product contains a pseudo-regular factor might not work. Also someone could try to find a counterexample that means a recognizable formal power series (over a commutative ring that is no integral domain), to prove that the iteration theorem does not hold. Indeed an approach might be to use non-trivial zero divisors in a tree such that the tree cannot be iterated.

References

[BR80] Jean Berstel and Christophe Reutenauer. "Recognizable Formal Power Series on Trees". In: *Theoretical Computer Science* (1980). DOI: 0304-3975/82/0000-0000/$02.75.

[Baz10] Bernard Bazioch. *Abstract Algebra I*. 2010. URL: http://www.math.buffalo.edu/~badzioch/MTH619/Lecture_Notes_files/MTH619_week13.pdf (visited on September 5, 2019).

[Bos13] Siegfried Bosch. *Algebra*. Springer Spektrum, 2013. DOI: 10.1007/978-3-642-39567-3.

[Fis13] Gerd Fischer. *Lineare Algebra*. Springer Spektrum, 2013. DOI: 10.1007/978-3-658-03945-5.

[Jac80] Gérard Jacob. "Un théorème de factorisation des produits d'endomorphismes de K^n". In: *Journal of Algebra* (1980). DOI: 10.1016/0021-8693(80)90080-0.

[KM17] Christian Karpfinger and Kurt Meyberg. *Algebra*. Springer Spektrum, 2017. DOI: 10.1007/978-3-662-54722-9.

[Lan04] Serge Lang. *Algebra*. Springer, 2004. DOI: 10.1007/978-1-4613-0041-0.

[Pau18] Alexander Paulin. *Introduction to Abstract Algebra*. 2018. URL: https://math.berkeley.edu/~apaulin/AbstractAlgebra.pdf (visited on September 5, 2019).

[Reu80] Christophe Reutenauer. "An Ogden-Like Iteration Lemma for Rational Power Series". In: *Acta Informatica* (1980). DOI: 10.1007/BF00263993.

[SG16] Hannes Stoppel and Birgit Griese. *Übungsbuch zur Linearen Algebra*. Springer Spektrum, 2016. DOI: 10.1007/978-3-658-14522-4.

[SS78] Arto Salomaa and Matti Soittola. *Automata-Theoretic Aspects of Formal Power Series*. Springer, 1978. DOI: 10.1007/978-1-4612-6264-0.

[Sti18] Jakob Stix. *Lineare Algebra*. 2018. URL: http://www.uni-frankfurt.de/74414804/Stix_LineareAlgebra_Skript.pdf (visited on August 14, 2019).

Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe, insbesondere sind wörtliche oder sinngemäße Zitate als solche gekennzeichnet. Mir ist bekannt, dass Zuwiderhandlung auch nachträglich zur Aberkennung des Abschlusses führen kann.
Ich versichere, dass das elektronische Exemplar mit den gedruckten Exemplaren übereinstimmt.

Leipzig, den 24.09.2019

.....
Patrick Kramer