Sisältö

1. Modulit

Vektoriavaruudet ovat vaihdannaisia ryhmiä, joissa on määritelty jonkin kunnan skalaaritoiminta. Hyväksymällä kerroinrakenteeksi kunnan sijaan rengas saadaan rakenne nimeltä *moduli*. Modulin käsite on siis vektoriavaruuden yleistys, mutta modulien teoria poikkeaa melko paljon vektoriavaruuksien teoriasta. Yleisellä modulilla ei esimerkiksi välttämättä ole kantaa, ja vaikka olisikin, kannan pituus ei ole välttämättä yksikäsitteinen, jolloin dimension käsitettä ei voida määritellä. Toisaalta jokaiselle vaihdannaiselle ryhmälle voidaan määritellä luonnollinen modulirakenne, missä alkioita kerrotaan kokonaisluvuilla, mistä johtuen modulien teoria on myös suurelta osin vaihdannaisten ryhmien teoriaa.

1.1. Modulit ja lineaarikuvaukset.

DEFINITION 1.1. Olkoon R rengas, ei välttämättä vaihdannainen. Vaihdannaista ryhmää $(M, +)$, jossa on määritelty renkaan R lineaarinen toiminta, nimitetään *moduliksi*. Renkaan lineaarinen toiminta on renkaan kertolaskumonoidin $(R, \cdot)$ toiminta, joka toteuttaa seuraavat ehdot kaikilla $a, b \in R$ ja $x, y \in M$:

(M1) $1.x = x$
(M2) $(ab).x = a.(b.x)$
(M3) $(a + b).x = a.x + b.x$
(M4) $a.(x + y) = a.x + a.y$.

Rengasta R kutsutaan modulin *kerroinrenkaaksi*, ja sen toimintaa *skalaarikertolaskuksi*.

Modulia, jossa kerroinrenkaana on R, voidaan nimittää R-moduliksi. Aksioomat (M1) ja (M2) määrittelevät renkaan kertolaskumonoidin toiminnan, aksiooma (M3) kertoo, miten renkaan yhteenlasku suhtautuu tähän toimintaan, ja aksiooma (M4) varmistaa, että toiminta on lineaarista (vrt. lineaarikuvauksiin). Modulin aksioomista voidaan helposti johtaa tuttuja laskusääntöjä, kuten $0.x = 0$, $(-1).x = -x$ jne. Yleensä toimintaa merkitään yksinkertaisesti kertolaskuna jättämällä alkioiden välistä piste pois.

Huomaa, että renkaan toiminnan olemassaolo pitää sisällään sen oletuksen, että $a.x \in M$ kaikilla $a \in R$ ja $x \in M$. Tämä voidaan myös ilmaista sanomalla, että modulin täytyy olla *suljettu skalaarikertolaskun suhteen*.

Tässä määritelty renkaan toiminta on tarkasti ottaen renkaan *vasen* toiminta, ja siksi tällaista modulia nimitetään joskus *vasemmaksi* R-moduliksi. Vastaavasti voitaisiin määritellä oikeat modulit renkaan oikean toiminnan avulla.

Esimerkkejä moduleista:

- Jos K on kunta, jokainen K-vektoriavaruus on samalla K-moduli, sillä modulin aksioomat ovat tällöin täsmälleen samat kuin vektoriavaruuden aksioomat.
- Rengas R on itse R-moduli, kun skalaarikertolaskuksi otetaan renkaan oma kertolasku.
- Jokainen vaihdannainen ryhmä on $\mathbb{Z}$-moduli, kun skalaarikertolaskuksi määritellään monikerran ottaminen: $n.x = nx = x + \cdots + x$ (n kertaa). Tämä on itse asiassa ainoa tapa, jolla $\mathbb{Z}$ voi toimia vaihdannaisessa

ryhmässä, sillä renkaan $\mathbb{Z}$ additiivinen ryhmä on alkion 1 virittämä, ja toiminta määräytyy tällöin täysin aksioomista (M1) ja (M3).

- Jäännösluokkarenkaiden $\mathbb{Z}_n$ toiminta ryhmässä M on myös yksikäsitteisesti määrätty: $[k]_n.x = kx$ (monikerta). Jotta tällainen toiminta olisi hyvin määritelty, täytyy ryhmässä M päteä $nx = 0$, eli jokaisen alkion kertaluvun täytyy jakaa luku n. Tämä toteutuu muun muassa silloin kun $|M| = n$. Kuitenkin esimerkiksi Kleinin neliryhmä on $\mathbb{Z}_2$-moduli. Kun p on alkuluku, rengas $\mathbb{Z}_p$ on kunta, ja jokainen $\mathbb{Z}_p$-moduli on siis vektoriavaruus.
- Olkoon K kunta. Kaikki K-kertoimiset $n \times n$ -matriisit muodostavat renkaan $M_n(K)$, joka ei ole vaihdannainen. Tämä rengas toimii matriisikertolaskulla vasemmalta sarakevektorien avaruudessa K^n ja oikealta vastaavassa rivivektorien avaruudessa. Vektoriavaruutta K^n voidaan siis tarkastella joko vasempana tai oikeana $M_n(K)$-modulina. Nämä kaksi struktuuria ovat lisäksi täysin samanlaiset.
- Renkaan R ideaalit ovat R-moduleja, kun kertolaskuna on renkaan oma kertolasku. Ideaalit ovat samalla rengasmodulin R alimoduleja (määritelmä seuraa). Alirenkaat sen sijaan eivät yleensä ole alimoduleja, koska ne eivät ole vakaita renkaan kertolaskutoiminnassa.

Olkoot M ja N joitain R-moduleja. Kuvausta $f\colon M \to N$ kutsutaan *R-modulihomomorfismiksi* tai *R-lineaarikuvaukseksi*, jos se on skalaarikertolaskun säilyttävä ryhmähomomorfismi, eli seuraavat ehdot pätevät kaikilla $x, y \in M$ ja $a \in R$:

(L1) $f(x + y) = f(x) + f(y)$
(L2) $f(a.x) = a.f(x)$.

Bijektiivistä lineaarikuvausta nimitetään *lineaariseksi isomorfismiksi*. Lineaarikuvauksen ydin on sama kuin vastaavan ryhmähomomorfismin ydin, eli nollan alkukuva.

Lineaarisuusehdot voidaan myös yhdistää yhdeksi *lineaarisuuskriteeriksi*, joka on toisinaan kätevämpi tarkistaa:

(LK) $f(a.x + y) = a.f(x) + f(y)$ kaikilla $x, y \in M$ ja $a \in R$.

Example 1.2. Voidaan osoittaa, että jos rengas R on vaihdannainen, kaikkien R-modulihomomorfismien $M \to N$ joukko on itse R-moduli, kun laskutoimitukset määritellään pisteittäin:

$$(f + g)(x) = f(x) + g(x) \qquad \text{ja} \qquad (a.f)(x) = a.f(x).$$

Tätä modulia merkitään $\mathrm{Hom}_R(M, N)$, tai jos kerroinrengas on selvä asiayhteydestä, yksinkertaisemmin $\mathrm{Hom}(M, N)$. Tarkka todistus jätetään harjoitustehtäväksi. Huomaa, että ei ole edes itsestään selvää, että lineaarikuvausten $M \to N$ joukko on suljettu annettujen laskutoimitusten suhteen.

1.2. Ali- ja tekijämodulit. Modulin M *alimoduli* N on ryhmän M aliryhmä, joka on vakaa kertolaskutoiminnan suhteen. Kaikilla $x, y \in N$ ja $a \in R$ (kerroinrengas) täytyy siis päteä seuraavat ehdot:

(AM1) $N \neq \emptyset$
(AM2) $x - y \in N$
(AM3) $a.x \in N$.

Ehdot (AM1) ja (AM2) tulevat aliryhmäkriteeristä. Ehdoista (AM1) ja (AM3) seuraa, että $0_M \in N$.

Mielivaltaisten alimodulien leikkaus on aina alimoduli. Lineaarikuvausten kuvat ja ytimet ovat myös alimoduleja.

Olkoot A ja B kaksi modulin M alimodulia. Niiden *summa* on

$$A + B = \{a + b \mid a \in A, b \in B\}.$$

Tämä määritelmä on additiivinen versio aliryhmien tulon määritelmästä (katso luku **??**). Koska modulit ovat vaihdannaisia ryhmiä, alimodulien summa on aina aliryhmä. Se on samalla pienin aliryhmä, joka sisältää summattavansa, mikä voidaan ilmaista kaavalla $A + B = \langle A \cup B \rangle$. Lisäksi alimodulien summa on suljettu skalaarikertolaskun suhteen, koska $r(a+b) = ra+rb \in A+B$ pätee kaikilla $a \in A$ ja $b \in B$.

Summaa voidaan yleistää äärettömän monelle alimodulille yllä mainitun viritysominaisuuden avulla. Olkoon $(M_i)_{i\in I}$ perhe[1] modulin M alimoduleita. Määritellään näiden alimodulien summa seuraavasti:

$$\sum_{i\in I} M_i = \left\langle \bigcup_{i\in I} M_i \right\rangle.$$

Toisin sanoen summa on sellaisten alkioiden x virittämä aliryhmä, joista kukin sisältyy johonkin alimoduleista M_i. Summan alkiot ovat siis muotoa

$$x_{i_1} + x_{i_2} + \cdots + x_{i_n},$$

missä jokainen x_{i_k} sisältyy johonkin alimoduliin M_{i_k}. Tämä voidaan ilmaista myös sanomalla, että alkiot ovat summia $\sum_{i\in I} x_i$, missä $x_i \in M_i$ kaikilla i, ja $x_i = 0$ lukuunottamatta äärellistä määrää indeksejä i. Alimodulien yleinen summa on aina alimoduli.

Example 1.3. Tarkastellaan reaalilukujen yhteenlaskuryhmää $\mathbb{Z}$-modulina. Määritellään kullakin alkuluvulla p joukko

$$M_p = \{n/p^k \mid n \in \mathbb{Z}, k \in \mathbb{N}\}.$$

Joukot M_p ovat $\mathbb{Z}$-modulin $\mathbb{R}$ alimoduleja. Määritetään näiden alimodulien summa $S = \sum_p M_p$. Selvästikin jokaisella p pätee $M_p \subset \mathbb{Q}$, ja $\mathbb{Q}$ on modulin $\mathbb{R}$ alimoduli. Täten $S \subset \mathbb{Q}$, koska S on pienin alimoduli, joka sisältää kaikki modulit M_p. Toisaalta jokainen rationaaliluku voidaan ilmaista summana $\sum_{i=0}^{n} m_i/p_i^{k_i}$, missä osoittajat ovat kokonaislukuja ja nimittäjät alkulukujen potensseja. Siispä $S = \mathbb{Q}$.

Modulin M mikä tahansa alimoduli N on normaali aliryhmä, koska M on vaihdannainen ryhmä. Aliryhmän N suhteen voidaan siis muodostaa tekijäryhmä. Tästä tekijäryhmästä tulee samalla *tekijämoduli*, sillä sivuluokkien skalaarikertolasku

$$a(x + N) = ax + N$$

on automaattisesti hyvin määritelty. Jos nimittäin $x = y + n$ jollain $n \in N$, niin $ax = ay + an \in ay + N$, sillä $an \in N$. Tekijämodulia merkitään tavalliseen tapaan symbolilla M/N.

Tekijämoduleille pätee samanlainen homomorfialause kuin ryhmille ja renkaille. Lisäksi Noetherin isomorfialauseet pätevät myös modulien tapauksessa.

[1]Perheellä tarkoitetaan kuvausta $i \mapsto M_i$ indeksijoukolta I johonkin alimodulien joukkoon. Jos $I = \mathbb{N}$, tämä on sama kuin jono $(M_0, M_1, M_2, \dots)$.

1.3. Direct sums and products of modules. For most basic algebraic structures the product of two structures can be given the structure of same kind. (This doesn't apply to some special structures, such as integral domains or fields.) The product of two R-modules is called the *direct sum* and denoted by $M \oplus N$. It is an R-module, which consists of pairs (m, n), where $m \in M$ and $n \in N$. In the case of several modules, the sum is be denoted by

$$\bigoplus_{i=1}^{n} M_i,$$

and the elements are n-tuples $(m_1, m_2, \ldots, m_n)$, where $m_i \in M_i$ for every i. In the case of an infinite index set the definition of a direct sum differs from the definition of the cartesian product. Nevertheless, both are R-modules, and the latter is called the *direct product*.

DEFINITION 1.4. Let $(M_i)_{i\in I}$ be a family of R-modules. The *direct product* of the modules M_i consists of families of elements $x = (x_i)_{i\in I}$, where $x_i \in M_i$ for every i. The direct product is an R-module, when we define the operations pointwise:

$$(x + y)_i = x_i + y_i \qquad \text{and} \qquad (ax)_i = ax_i.$$

The direct product is denoted by $\prod_{i\in I} M_i$.

The direct sum is defined as a subset of the direct product. Assume again that $(M_i)_{i\in I}$ is a family of R-modules.

DEFINITION 1.5. The *direct sum* of modules M_i consists of families of elements $(x_i)_{i\in I}$, where $x_i \in M_i$ for every i and moreover $x_i \neq 0$ only for a finite number of indices. The direct sum is an R-module, when the operations are defined pointwise, as in the case of the direct product. The direct sum is denoted by $\bigoplus_{i\in I} M_i$.

The elements of the direct sum are thus families where only finitely many members differ from the zero element. Such a family is called *finitely supported*.

When considering the product and sum modules, we often refer to the so called *canonical projections* $\pi_j \colon \prod_{i\in I} M_i \to M_j$, $\pi_j(x) = x_j$ and *canonical injections* $\iota_j \colon M_j \to \bigoplus_{i\in I} M_i$, $\iota_j(y) = (x_i)_{i\in I}$, where

$$x_i = \begin{cases} y, & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

For example if the index set is $I = \{1, 2, 3, 4\}$ and $a \in M_2$, we may write $\iota_2(a) = (0, a, 0, 0)$. Both the canonical projections and canonical injections are module homomorphisms. Every element (x_i) of the direct sum can be written as a finite sum $\sum_i \iota_i(x_i)$. The same doesn't hold for the direct product, if the index set is infinite.

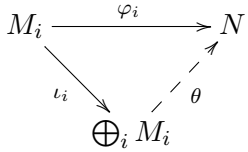
For the canonical injections the following theorem holds, it is called the *universal property of the direct sum*.

PROPOSITION 1.6. *Let (M_i) be a family of R-modules. Suppose also that N is an R-module and φ_i is an R-linear map $M_i \to N$ for every i. Then there exists a*

unique R-linear map $\theta\colon \bigoplus_i M_i \to N$, for which

$$\varphi_i = \theta \circ \iota_i \tag{1}$$

holds for every i, that is, the following diagram commutes:



Todistus. Every element $x = (x_i)$ of the direct sum can be written in the form $x = \sum_i \iota_i(x_i)$. Thus, if θ is linear and satisfies the condition (1), then for all $x \in \bigoplus_i M_i$ we must have

$$\theta(x) = \theta\left(\sum_i \iota_i(x_i)\right) = \sum_i (\theta \circ \iota_i)(x_i) = \sum_i \varphi_i(x_i).$$

This formula defines the values of the map θ uniquely.

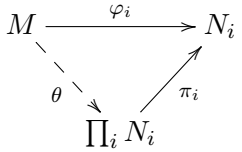
Then we prove that the map θ, defined by the above formula, satisfies the conditions mentioned in the theorem. It is easy to verify that θ is R-linear. Furthermore, if $y \in M_j$, then $(\iota_j(y))_i = 0$ for every $i \neq j$. Thus for every j we have

$$\theta(\iota_j(y)) = \sum_i \varphi_i((\iota_j(y))_i) = \varphi_j(y),$$

that is, the map θ satisfies the condition (1). □

By universality we mean the following: whenever we have a family of linear maps to some specific module, this family can be replaced by one map from the direct sum to the module in question. The direct sum is, in a way, a "universal" family of linear maps (ι_i), which can be completed with a linear map θ to correspond to any family of linear maps (φ_i). A theorem of similar nature holds also for the direct product and the canonical projections.

Proposition 1.7. *Let (N_i) be a family of R-modules. Suppose also that M is an R-module, and φ_i is an R-linear map $M \to N_i$ for every i. Then there exists a unique R-linear map $\theta\colon M \to \prod_i N_i$, for which $\varphi_i = \pi_i \circ \theta$ for every i, that is, the following diagram commutes.*

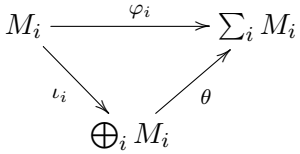


Todistus. Exercise. □

When we investigated groups, it was sometimes useful to know, if a certain group is isomorphic to some product group. Also for modules we have an analogous result.

Proposition 1.8. *Let $(M_i)_{i\in I}$ be a family of submodules of an R-module M. If $\sum_i M_i = M$ and $M_i \cap \sum_{j\neq i} M_j = \{0\}$ for every i, then M is isomorphic to the direct sum $\bigoplus_i M_i$.*

TODISTUS. For every i we can define the inclusion map $\varphi_i\colon M_i \to M$, where $\varphi_i(x) = x$. By the universal property of the direct sum we can find an R-linear map $\theta\colon \bigoplus_i M_i \to M$, for which $\theta(\iota_i(x)) = \varphi_i(x) = x$ for every i and every $x \in M_i$. We prove that θ is bijective.



First we notice that if $x = (x_i) \in \bigoplus_i M_i$, then

$$\theta(x) = \theta\left(\sum_i \iota_i(x_i)\right) = \sum_i \varphi_i(x_i) = \sum_i x_i.$$

To prove surjectivity, suppose that $y \in M$ is an arbitrary element. Since $M = \sum_i M_i$, the element y can be written as a finite sum $y = \sum_i x_i$, where $x_i \in M_i$ for every i. Now $x = \sum_i \iota_i(x_i)$ is an element of the direct sum $\bigoplus_i M_i$, and by the above we have that $\theta(x) = \sum_i x_i = y$.

Then suppose that $\theta(x) = \theta(y)$ for some $x, y \in \bigoplus_i M_i$. This means that $\sum_i x_i = \sum_i y_i$, that is, $\sum_i (x_i - y_i) = 0$. Moreover, for every i we have

$$(x_i - y_i) = -\sum_{i \neq j} (x_j - y_j).$$

The left hand side of the equation is an element of the submodule M_i, and the right hand side belongs to the sum module $\sum_{i \neq j} M_j$. By assumption their intersection is trivial, especially we have that $x_i - y_i = 0$. Since this holds for any i, we get $x_i = y_i$ for every i, thus $x = y$. This proves injectivity □

The above result clarifies, why direct sums of modules are in general more important in algebra than direct products. Think for example of the direct product module $\prod_{i \in I} \mathbb{R}$. If the index set is finite, then this is the ordinary vector space $\mathbb{R}^n$. In this space every coordinate axis is a submodule, which consists of elements, which are sequences of the form $\iota_i(x) = (0, \ldots, 0, x, 0, \ldots, 0)$. All elements of the space can be obtained by summing vectors of the coordinate axes. The coordinate axes are isomorphic with the module $\mathbb{R}$, and the above result gives the correspondence $\bigoplus_i \iota_i(\mathbb{R}) \cong \mathbb{R}^n$.

However, if the index set is infinite, there are elements in the space $\prod_{i \in I} \mathbb{R} = \mathbb{R}^I$, which *cannot* be obtained by summing vectors of the coordinate axes. The sum $\sum_i \iota_i(\mathbb{R})$ of the submodules is then a proper submodule, which by the above result corresponds to the fact that $\bigoplus_i \mathbb{R}$ is a proper subset of the direct product $\mathbb{R}^I$.

Remark. In group theory the direct sum of commutative groups $(G_i, +)$ is constructed similarly as the direct sum of modules. However, the exactly same construction is called the direct product in the case that the group operation is denoted by multiplication. In both structures the elements are families (g_i), where g_i is the neutral element for all but a finite number of indices. If we give up this requirement of finiteness, we get a structure, which corresponds to the direct product of modules, which in the case of groups is called the *unlimited* direct product or sum, depending on the notation of the binary operation.

	modules	groups
direct sum	finitely supported product	finitely supported, operation denoted as addition
direct product	cartesian product	finitely supported, operation denoted as multiplication
unlimited product	–	cartesian product

TAULUKKO 1. Differences between the namings of sums and products for modules and groups

2. Constructions with modules

2.1. Free modules. One of the most important features of vector spaces is that the vectors can be expressed as linear combinations of basis vectors in a unique way. In this chapter we consider modules with the corresponding property.

Let X be a set of elements of an R-module M. Fix some indexing $X = \{x_i\}_{i\in I}$ of the set X. A finite sum $\sum_i r_i x_i$, where $r_i \in R$ for every i, is called a *linear combination* of the set X. If every element of the module M can be expressed as a linear combination of elements of the set X, we say that X *generates* or *spans* the module M. Furthermore, if for every linear combination we have that $\sum_i r_i x_i = 0$ only, when $r_i = 0$ for every i, we say that the subset X is *linearly independent* or *free*.

DEFINITION 2.1. Let M be an R-module. A subset $B \subset M$ is called a *basis* of the module M, if B generates the module M and it is linearly independent. If such a subset exists, the module M is called *free*.

In a free module every element can be written as a linear combination of basis elements. Moreover, this representation is unique: if $\sum_i r_i b_i = \sum_i r'_i b_i$, then $\sum_i (r_i - r'_i) b_i = 0$, and because B is free, it follows that $r_i = r'_i$ for all i.

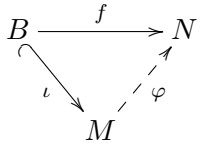
Remark. When considering linear independence, it is essential that the set of indices is chosen, and that same indices don't appear several times in a linear combination. Otherwise one could say that the singleton set $\{x_1\}$ is not free, because the linear combination $x_1 - x_1$ is zero, even though the coefficients are not zero. Sometimes we, instead of the set X, use an indexed sequence or a family $(x_i)_{i\in I}$. The difference becomes visible in situations where the same element is repeated; for example the sequence (x, x) is not free, but thought of as a set $\{x, x\} = \{x\}$ is free.

Examples of free modules:

- In a basic course on linear algebra, it is proved that every finitely generated $\mathbb{R}$-vector space has a basis, thus every such vector space is a free $\mathbb{R}$-module. The same proof works for any coefficient field. Also other than finitely generated vector spaces have a basis, but the proof requires Zorn's lemma.
- Any ring R is a free R-module, with the singleton $\{1\}$ as basis. More generally: the product module R^n is free, and the *natural basis* consists of elements $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ (where the unit element is at the ith place).
- If R is a ring, the set $R^{n\times m}$ of $n \times m$ -matrices with coefficients in R, is an R-module, with addition and multiplication of matrices as binary operations. The natural basis consists of the elementary matrices E_{ij}, where the element on row i and column j is the unit element of the ring R, all other elements are zero.
- A commutative group is called free, if it is free as a $\mathbb{Z}$-module. The group $\mathbb{Q}$ is not a free group. Also the group $\mathbb{Z}_n$ is not free, which for example follows from Proposition 2.3, which we shall soon prove. This result shows that every free commutative group is isomorphic to a direct sum consisting of groups $\mathbb{Z}$; especially every free group is infinite.

The free modules have the following universal property. It claims for example that every linear map defined on a free module is determined by the images of the basis elements.

PROPOSITION 2.2. *Let M be a free R-module, with basis B, and denote the inclusion map by $\iota\colon B \to M$. Suppose also that N is another R-module and $f\colon B \to N$ is any map. Then there exists a unique R-linear map $\varphi\colon M \to N$, for which $\varphi(b) = f(b)$ for all $b \in B$, that is, the following diagram commutes.*



Moreover

i) *φ is injective, if and only if f is injective and the image set fB is free*
ii) *φ is surjective, if and only if the image set fB generates the module N.*

TODISTUS. Every element x of the free module M has a unique representation $x = \sum_i r_i b_i$ as a linear conbination of basis elements. If $\varphi\colon M \to N$ is a linear map, for which $\varphi(b) = f(b)$ for all $b \in B$, then

$$\varphi(x) = \varphi\left(\sum_i r_i b_i\right) = \sum_i r_i \varphi(b_i) = \sum_i r_i f(b_i). \tag{2}$$

The linear map, which we are looking for, has to satisfy the above condition for every x, thus the map is unique, if it exists.

On the other hand, we can use the condition (2) to define a map $\varphi\colon M \to N$. Now it is easy to verify that the map defined by the formula $\sum_i r_i b_i \mapsto \sum_i r_i f(b_i)$ is R-linear, and that we have $b \mapsto f(b)$ for every basis element b. Thus the map satisfies the desired conditions.

We leave the proofs of the two last claims as an exercise. □

When a ring R is thought of as an R-module, we can construct the direct sum $\bigoplus_{i\in I} R$, which is denoted by $R^{(I)}$. This is a free module. Its' *natural basis* consists of elements $e_j = (\delta_{ij})_{i\in I}$, where $j \in I$ and

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

The elements of the basis are thus images of the neutral element in the canonical injections $\iota_i\colon R \to \bigoplus_i R$. Next we prove that every free module is isomorphic to such a direct sum.

PROPOSITION 2.3. *If M is a free R-module, then $M \cong R^{(I)}$ for some I.*

TODISTUS. Let $B = \{b_i\}_{i\in I}$ be a basis of the free module M. Define a map $f\colon B \to R^{(I)}$ by the formula $f(b_i) = e_i$. By the universal property 2.2 there exists an R-linear map $\varphi\colon M \to R^{(I)}$, for which $\varphi(b_i) = e_i$ for all i. By the latter part of the same theorem we have that φ is bijective, since the set $\{e_i\}$ is a basis for the module $R^{(I)}$. □

Now the universal property of a free module can be interperted as follows: For every set I we can construct a "universal" R-module $R^{(I)}$. In this module

the element i of the set I is often identified with the element e_i of the natural basis. Then every map f from the set I to an R-module N can be extended to a homomorphism $\varphi\colon R^{(I)} \to N$.

EXAMPLE 2.4. Let R be a ring and X any set. Identify each element $x \in X$ with the basis element e_x of the free module $R^{(X)}$. Then an arbitrary element of the free module can be written as a formal linear combination

$$\sum_x r_x x = \sum_x r_x e_x.$$

In this way we can construct a module structure over any set X; this is called the *free module generated by the set* X.

Especially, if $R = \mathbb{Z}$ and $n \in R$, the element nx can be considered as a formal multiple of x. In this way we obtain the free commutative group generated by the set X. In this group the elements of the set X can for example be added together.

In algebraic topology we encounter a situation, where the set X consists of generalized triangles or *simplexes* of some topological space T. (More precisely, these are images of euclidean triangles and their n-dimensional analogues, such as tetrahedra, in continuous maps.) In the free group $\mathbb{Z}^{(X)}$ we can form linear combinations of these simplexes, and by investigating this group, obtain information about the space T.

2.2. Tensor products. Several products defined in vector spaces are linear with respect to both components, in which case we call them *bilinear*. If the product of vectors is denoted by $(x, y) \mapsto x \otimes y$, bilinearity means thus that

$$\begin{aligned} &(x+y)\otimes z = x\otimes z + y\otimes z, && (ax)\otimes y = a(x\otimes y) \\ \text{and}\quad &x\otimes(y+z) = x\otimes y + x\otimes z, && x\otimes(ay) = a(x\otimes y). \end{aligned}$$

For example the ordinary dot product $x \cdot y$ and the cross product $x \times y$ in three dimensional space are bilinear products. In the following we generalize the concept of a bilinear product to arbitrary modules, and investigate the *tensor product* of modules, in which one can define a "universal" bilinear product.

DEFINITION 2.5. Let R be a commutative ring, and let M, N and P be R-modules. A map f from the set $M \times N$ to the module P is called *R-bilinear*, if it is linear with respect to both components, that is, for every $x, y \in M$, $z, w \in N$ and $a \in R$ the following holds:

(B1) $f(x+y, z) = f(x, z) + f(y, z)$
(B2) $f(x, z+w) = f(x, z) + f(x, w)$
(B3) $f(ax, z) = af(x, z)$
(B4) $f(x, az) = af(x, z)$.

For example the dot product is a bilinear map $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$. In general, the modules M and N need not be the same. Observe, that for a bilinear map the following formulas hold: $f(x, 0) = 0$ and $f(0, y) = 0$, because for example $f(x, 0) = f(x, 0.0) = 0.f(x, 0) = 0$.

Let M and N be arbitrary R-modules. We construct a module T, for which one can define a bilinear map $M \times N \to T$, $(x, y) \mapsto x \otimes y$. The idea is to start with a module, whose elements are linear combinations of pairs (x, y). These pairs may be regarded as formal products. After that we identify elements in such a

way that the bilinearity conditions are satisfied: for example every pair $(x+y,z)$ is identified with the linear combination $(x,z)+(y,z)$.

Let C be the free R-module $R^{(M\times N)}$. The natural basis for this module consists of the families $e_{(x,y)}$, where $(x,y)\in M\times N$. As before, we identify every basis element with the corresponding pair (x,y). Then C consists of linear combinations of such pairs, with coefficients in the ring R. Consider the following four types of linear combinations, where $x,y\in M$, $z,w\in N$ and $a\in R$:

$$\begin{aligned}&(x+y,z)-(x,z)-(y,z)\\&(x,z+w)-(x,z)-(x,w)\\&(ax,z)-a(x,z)\\&(x,az)-a(x,z).\end{aligned}$$

Let D be the submodule of C, which is generated by the linear combinations of the above mentioned forms.

DEFINITION 2.6. The *tensor product* $M\otimes_R N$ of two R-modules M and N is the quotient module C/D, where $C=R^{(M\times N)}$ and D is the submodule defined above. If the coefficient ring is clear from the context, we can also use the simpler notation $M\otimes N$.

The equivalence class of an element (x,y) in the quotient module C/D is denoted by $x\otimes y$. (To be exact, this is the equivalence class of the element $e_{(x,y)}$, but this family was identified with the pair (x,y).) Because the pairs (x,y) generate the free module C, their equivalence classes generate the module C/D. Thus every element of the tensor product $M\otimes N$ may be written as a linear combination of some elements $x\otimes y$. The canonical map of the tensor product is the map $\eta\colon M\times N\to M\otimes N$, for which $\eta(x,y)=x\otimes y$. The canonical map is R-bilinear.

In the tensor product, the linear combinations of the generators $x\otimes y$ can be presented as sums without scalar coefficients. Namely, if $r_i\in R$, $x_i\in M$ and $y_i\in N$ for every i, then

$$\sum_i r_i(x_i\otimes y_i)=\sum_i (r_ix_i)\otimes y_i,$$

and $r_ix_i\in M$ for every i. An arbitrary element of the tensor product can thus be written in the form $\sum_i x_i\otimes y_i$.

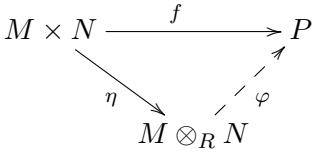
EXAMPLE 2.7. Suppose that m and n are relatively prime (g.c.d.=1) natural numbers, and consider the tensor product of the $\mathbb{Z}$-modules $\mathbb{Z}_m$ and $\mathbb{Z}_n$. Since m and n are relatively prime, there exist integers a and b, for which $am+bn=1$. Then for the element $\overline{x}\otimes\overline{y}\in\mathbb{Z}_m\otimes_{\mathbb{Z}}\mathbb{Z}_n$ the following holds:

$$\begin{aligned}\overline{x}\otimes\overline{y}&=(am+bn).(\overline{x}\otimes\overline{y})=am.(\overline{x}\otimes\overline{y})+bn.(\overline{x}\otimes\overline{y})\\&=a.(m\overline{x}\otimes\overline{y})+b.(\overline{x}\otimes n\overline{y})=a.(\overline{0}\otimes\overline{y})+b.(\overline{x}\otimes\overline{0})=0.\end{aligned}$$

Since every generator of the module $\mathbb{Z}_m\otimes_{\mathbb{Z}}\mathbb{Z}_n$ is zero, the tensor product is the trivial module.

When proving properties concerning the tensor product, it is usually not necessary to refer to the definition, but one can use the following universal property.

PROPOSITION 2.8. *Let M, N and P be R-modules, and let $f\colon M \times N \to P$ be an R-bilinear map to an R-module P. Then there exists a unique R-linear map $\varphi\colon M \otimes_R N \to P$, for which $\varphi(x \otimes y) = f(x, y)$ for every $x \in M$ and $y \in N$, that is, the following diagram commutes.*

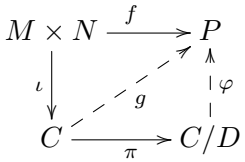


TODISTUS. Since the pairs (x, y), where $x \in M$ and $y \in N$, form a basis for the free module $C = R^{(M\times N)}$, then f can in a unique way be extended to a linear map $g\colon C \to P$ by the universal property of the free module. Let u be some generating element of the submodule D defined in the construction of the tensor product. Then $g(u) = 0$, since f is bilinear: for example, if $u = (ax, y) - a(x, y)$, then

$$g(u) = g\big((ax, y) - a(x, y)\big) = g(ax, y) - ag(x, y) = f(ax, y) - af(x, y) = 0.$$

Since $g(u) = 0$ for every generator u of D, we see that $D \subset \operatorname{Ker} g$. Thus there exists a unique R-module homomorphism $\varphi\colon C/D \to P$, for which $g = \varphi \circ \pi$, where π is the canonical surjection. Now for all $(x, y) \in M \times N$ we have $x \otimes y = \pi(x, y)$, thus

$$\varphi(x \otimes y) = \varphi(\pi(x, y)) = g(x, y) = f(x, y). \qquad \square$$



KUVA 1. The commutative diagram related to the proof of Theorem 2.8. The upper triangle is obtained from the universal property of the free module, and the lower triangle from the homomorphism theorem for modules. The map ι is the inclusion map, $\eta = \pi \circ \iota$.

EXAMPLE 2.9. Let m and n be relatively prime natural numbers. By the universal property, there doesn't exist a non-trivial $\mathbb{Z}$-bilinear map from the set $\mathbb{Z}_m \times \mathbb{Z}_n$ to any module P. Namely, if $f\colon \mathbb{Z}_m \times \mathbb{Z}_n \to P$ is bilinear, there would exist a linear map $\varphi\colon \mathbb{Z}_m \otimes_{\mathbb{Z}} \mathbb{Z}_n \to P$, for which $\varphi \circ \eta = f$. Previously we saw that $\mathbb{Z}_m \otimes_{\mathbb{Z}} \mathbb{Z}_n = 0$, so the map φ has to be the zero map. From this it follows that also f is the zero map.

EXAMPLE 2.10. Let R be any (commutative) ring. Consider the free product modules R^m and R^n. (If R is a field, then these are finite dimensional vector spaces.) The tensor product of these modules has a simple concrete interpretation.

Denote by the symbol $R^{m\times n}$ the set of all $m \times n$ -matrices, with coefficients in R. These matrices form a free R-module. The natural basis consists of the elementary matrices E_{ij}, where the element on row i and column j is the unit

element, all other elements are zero. The *outer product* or *dyadic product* $g(x,y)$ of the elements $x=(x_1,\dots,x_m)$ and $y=(y_1,\dots,y_n)$ is defined to be the matrix

$$g(x,y)=\begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & & & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix}.$$

The map $g\colon R^m\times R^n\to R^{m\times n}$ is R-bilinear.

We prove that $R^m\otimes R^n\cong R^{m\times n}$ by using the universal property of the tensor product. Since the map g is bilinear, there exists an R-linear map $\varphi\colon R^m\otimes R^n\to R^{m\times n}$, for which $\varphi\circ\eta=g$. We prove that φ is bijective.

Every element of the module $R^{m\times n}$ can be written as a linear combination of elementary matrices $\sum_{i,j}a_{ij}E_{ij}$. When we denote the natural bases of the modules R^m and R^n by $\{e_i\}_{i=1}^m$ and $\{e_j\}_{j=1}^n$, we see that

$$\varphi\left(\sum_{i,j}a_{ij}(e_i\otimes e_j)\right)=\sum_{i,j}a_{ij}\varphi(\eta(e_i,e_j))=\sum_{i,j}a_{ij}g(e_i,e_j)=\sum_{i,j}a_{ij}E_{ij}.$$

Thus φ is surjective. Let then $\sum_{i,j}a_{ij}(e_i\otimes e_j)\in\operatorname{Ker}\varphi$. Then

$$0=\varphi\left(\sum_{i,j}a_{ij}(e_i\otimes e_j)\right)=\sum_{i,j}a_{ij}E_{ij},$$

and because the set $\{E_{ij}\}_{i,j}$ is free, we must have $a_{ij}=0$ for all i,j. Thus $\operatorname{Ker}\varphi=\{0\}$, and the map φ is injective.

Now we see that the ordinary dot and cross products of the space $\mathbb{R}^n$ can be derived from the universal tensor product. The dot product $x\cdot y$ comes from the matrix $A=x\otimes y$ by the linear map $A\mapsto\sum_i A_{ii}$, which adds up all diagonal elements. The cross product of the three-dimensional space is obtained by the map

$$A\mapsto(A_{23}-A_{32},\,A_{31}-A_{13},\,A_{12}-A_{21}),$$

which also is $\mathbb{R}$-linear.

In the following we list some properties of the tensor product.

PROPOSITION 2.11. *Let M, N and P be R-modules. Then there exist the following unique R-module isomorphisms:*

i)	$M\otimes N\cong N\otimes M$,	*where* $x\otimes y\mapsto y\otimes x$
ii)	$(M\otimes N)\otimes P\cong M\otimes(N\otimes P)$,	*where* $(x\otimes y)\otimes z\mapsto x\otimes(y\otimes x)$
iii)	$(M\oplus N)\otimes P\cong(M\otimes P)\oplus(N\otimes P)$,	*where* $(x,y)\otimes z\mapsto(x\otimes z,y\otimes z)$
iv)	$R\otimes M\cong M$,	*where* $a\otimes x\mapsto a.x$.

In the last claim we think of R as an R-module.

TODISTUS. We prove item (i) and leave the others as exercises. Define the maps $f\colon M\times N\to N\otimes M$ and $g\colon N\times M\to M\otimes N$ by the formulas

$$f(x,y)=y\otimes x\qquad\text{and}\qquad g(y,x)=x\otimes y.$$

These maps are clearly bilinear, so by Proposition 2.8 there exist unique linear maps $\varphi\colon M\otimes N\to N\otimes M$ and $\psi\colon N\otimes M\to M\otimes N$, for which $\varphi(x\otimes y)=f(x,y)=y\otimes x$ and $\varphi(y\otimes x)=g(y,x)=x\otimes y$ for all $x\in M$ and $y\in N$. Moreover

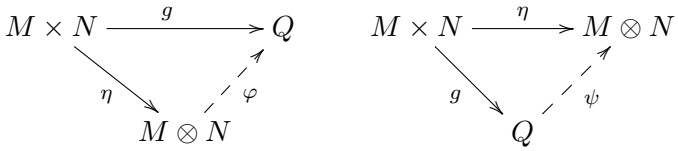
$\varphi \circ \psi = \mathrm{id}$ and $\psi \circ \varphi = \mathrm{id}$, thus φ and ψ are inverse functions of each other, and thus isomorphisms. □

Remark. The previous proposition may be interpreted that in a way the R-modules themselves form an algebraic structure, whose operations $\oplus$ and $\otimes$ satisfy the conditions (i)–(iv) mentioned in the Proposition. The "neutral element" of the multiplication $\otimes$ is the coefficient ring R itself.

2.3. Further information: characterization of the tensor product. It turns out that the universal property completely characterizes the tensor product (up to isomorphism). Because of this, the universal property can be used in several examples and proofs, instead of using the original definition.

PROPOSITION 2.12. *Let M, N and Q be R-modules, and let g be an R-bilinear map $M \times N \to Q$. Suppose that* $\operatorname{Im} g$ *generates the module Q and that the following condition holds: If f is any R-bilinear map from the product $M \times N$ to a module P, then there exists an R-linear map $\phi\colon Q \to P$, such that $f = \phi \circ g$. Then $Q \cong M \otimes_R N$, and for this isomorphism it holds that $g(x, y) \mapsto x \otimes y$.*

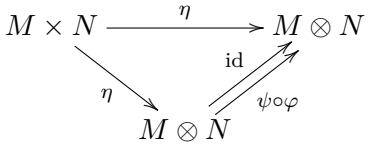
TODISTUS. Since the map g is bilinear, by the universal property of the tensor product, there exists a linar map $\varphi\colon M \otimes N \to Q$, for which $\varphi \circ \eta = g$. On the other hand, the canonical map η is bilinear, so by the assumption there exists a linear map $\psi\colon Q \to M \otimes N$, for which $\psi \circ g = \eta$.



We prove that φ is the inverse function of ψ. Clearly $\mathrm{id} = \mathrm{id}_{M \otimes N}$ is an R-linear map, for which $\mathrm{id} \circ \eta = \eta$. On the other hand, also $\psi \circ \varphi$ is R-linear, and

$$(\psi \circ \varphi) \circ \eta = \psi \circ g = \eta.$$

By the universal property of the tensor product, there can exist only one such map, thus $\mathrm{id} = \psi \circ \varphi$.



Then let $u \in Q$ be arbitrary. Since the image $\operatorname{Im} g$ generates the module Q, we can write $u = \sum_i g(x_i, y_i)$ for some $x_i \in M$ and $y_i \in N$. On the other hand, we see that

$$(\varphi \circ \psi) \circ g = \varphi \circ \eta = g,$$

hence

$$(\varphi \circ \psi)(u) = \sum_i (\varphi \circ \psi)(g(x_i, y_i)) = \sum_i g(x_i, y_i) = u.$$

Thus $\varphi \circ \psi = \mathrm{id}$, and φ is the inverse function of ψ. Thus $\psi\colon Q \to M \otimes N$ is an isomorphism of R-modules, for which it holds, that $\psi(g(x, y)) = \eta(x, y) = x \otimes y$. □

EXAMPLE 2.13. We prove the isomorphism of Example 2.10 by using Proposition 2.12. Let $g\colon R^m \times R^n \to R^{m\times n}$ be the bilinear map of that example. Clearly $\operatorname{Im} g$ generates the module $R^{m\times n}$, because $g(e_i, e_j) = E_{ij}$ for all i, j. Let now $f\colon R^m \times R^n \to P$ be any bilinear map to some R-module P. Our aim is to show that there exists an R-linear map $\phi\colon R^{m\times n} \to P$, for which $\phi \circ g = f$.

Define first the map ϕ from the set of the elementary matrices to the module P:

$$\phi(E_{ij}) = f(e_i, e_j).$$

By the universal property of the free module, ϕ can be extended in a unique way to a linear map from the whole module $R^{m\times n}$. If $A = (a_{ij}) \in R^{m\times n}$, then

$$\phi(A) = \phi\left(\sum_{i,j} a_{ij}E_{ij}\right) = \sum_{i,j} a_{ij} f(e_i, e_j).$$

Hence

$$\phi(g(x, y)) = \sum_{i,j} x_i y_j f(e_i, e_j) = f\left(\sum_i x_i e_i, \sum_j y_j e_j\right) = f(x, y)$$

for all $(x, y) \in R^m \times R^n$. Thus $f = \phi \circ g$, and from Proposition 2.12 we obtain that $R^{m\times n} \cong R^m \otimes R^n$. The dyadic product $g(x, y)$ and the tensor product $x \otimes y$ correspond to each other.

2.4. Further information: extension of scalars. We consider one application of tensor products, which is called *extension of scalars*. Let R and S be rings, and let $f\colon R \to S$ be a ring homomorphism. Now the ring S can be thought of as an R-module, when we define the scalar product by the formula $a.b = f(a) \cdot b$. If M is any R-module, then we can construct the tensor product

$$M_S = S \otimes_R M.$$

This tensor product is an S-module, when we define $b'.(b \otimes x) = (b'b) \otimes x$ for all elements $b, b' \in S$ and $x \in M$. We say that M_S is obtained from M by *extension of scalars*. Often R is in fact a subring of S, and $f\colon R \to S$ is the inclusion map. If $S = R$ and f is the identity map, then $M_R \cong M$ by Proposition 2.11.

LEMMA 2.14. *If $M = R^n$, where $n \in \mathbb{N}$, then M_S and S^n are isomorphic as S-modules.*

TODISTUS. Exercise. □

PROPOSITION 2.15. *Suppose that R is an integral domain. If R^m and R^n are isomorphic R-modules, then $m = n$.*

TODISTUS. The idea is to extend the scalar ring to a field and then use the concept of dimension. Denote $M = R^m$ and $N = R^n$ and suppose that $M \cong N$. Let K be the division field of the ring R. The field K becomes an R-module through the canonical map $\eta\colon R \to K$, $\eta(a) = a/1$. By the previous lemma

$$K^m \cong M_K \cong N_K \cong K^n.$$

The above isomorphisms are isomorphisms between K-vector spaces. Because the dimension of a vector space is unique, we have that $m = n$. □

The previous proposition holds also, when R is any commutative ring, not necessarily an integral domain. The proof is otherwise similar, but there the extension of scalars is done using a different map. If R is not an integral domain, it cannot be embedded into a field. Instead one can find a maximal ideal and form the quotient ring, which will be a field. The canonical surjection then makes the quotient ring an R-module.

If the ring R is non-commutative, it is possible that the modules R^m and R^n are isomorphic as right and left R-modules, but nevertheless $m \neq n$. The extension of scalars cannot be applied in this situation, since the tensor product is defined only when R is commutative.

3. Algebras

In several applications one deals with modules, in which there is also defined an internal bilinear multiplication. For example matrices can be multiplied with each other, and the multiplication behaves well with respect to the addition and scalar multiplication. This structure is called an *algebra*. In different books the definition might include more conditions for the multiplication, such as associativity or existence of a neutral element. In this material we keep the basic requirements in a minimum.

3.1. Basic properties. We start with the definition of an algebra.

DEFINITION 3.1. Let R be a commutative ring, and let A be an R-module, where there is defined an R-bilinear multiplication $(x, y) \mapsto x \cdot y$ for all $x, y \in A$. This kind of module A is called an *R-algebra*. If the multiplication is associative or commutative or if it has a neutral element, the algebra is called *associative*, *commutative* or *unitary*, respectively.

An algebra has thus three operations: addition, multiplication and scalar multiplication. The addition is a group operation, the distributive laws hold for both multiplications, and the scalar coefficients go inside the sums and products. Usually the multiplication is denoted simply by xy, and leave the dot out. Also the scalar multiplication can be denoted simply $a.x = ax$. If there is a danger of confusion, different notation may be used. For example the following laws hold in every algebra:

$$\begin{aligned} (a+b)x &= ax + bx & a(x \cdot y) &= (ax) \cdot y = x \cdot (ay) \\ (x+y) \cdot z &= x \cdot z + y \cdot z & -(x \cdot y) &= (-x) \cdot y = x \cdot (-y) \\ a(x+y) &= ax + ay & 0_R.x &= 0_A \cdot x = x \cdot 0_A = 0_A. \\ (-1).x &= -x \end{aligned}$$

The multiplication of an associative unitary algebra satisfies the conditions of multiplication in a ring, so an algebra can also be regarded as a ring (not necessarily commutative), where there is also a scalar multiplication. On the other hand, every commutative ring R is an R-module with respect to its' internal multiplication, so a commutative ring R is an associative commutative unitary R-algebra.

KUVA 2. An algebra is a module M, where there is defined a bilinear multiplication. An associative unitary algebra can also be thought of as a ring S, where there is defined a scalar multiplication of a (possibly different) ring.

Examples of algebras:

- Let R be a ring. In the module $R^{n \times n}$ of square matrices we can define the familiar matrix multiplication, which makes this module the *matrix algebra*.
- In the polynomial ring $R[X_1, \ldots, X_n]$ the coefficient ring R can be identified with the set of the constant polynomials. Then scalar multiplication can be defined with the same rule as multiplication of polynomials, and the polynomial ring becomes the commutative *polynomial algebra*.

- Let R be any ring, not necessarily commutative. As with groups, R can be equipped with scalar multiplication of the ring $\mathbb{Z}$: $n.a = a + \cdots + a$ (n times). Thus every ring is a $\mathbb{Z}$-algebra. As with groups, this is the only way, which $\mathbb{Z}$ can act in the ring R, thus the theory of $\mathbb{Z}$-algebras corresponds to the theory of rings.
- As we noticed above, any commutative ring R is an R-algebra. More generally, if R and S are commutative rings and $f\colon R \to S$ is a ring homomorphism, then S can be equipped with the scalar multiplication $a.b = f(a) \cdot b$. Then the ring S becomes an R-algebra.
- If K is a field, then every K-algebra A is a vector space. Then we can for example talk about the *dimension* of an algebra. If we furthermore have some extra structure in the vector space, such as a norm or a topology, we can talk about normed or topological algebras, respectively.
- As a vector space, the field of complex numbers $\mathbb{C}$ can be identified with the plane $\mathbb{R}^2$. The multiplication of complex numbers behaves well with respect to the vector operations of $\mathbb{R}^2$, and thus $\mathbb{C}$ is a two dimensional $\mathbb{R}$-algebra.
- Let M be an R-module. The module $\mathrm{End}_R(M) = \mathrm{Hom}_R(M, M)$ consisting of linear maps $M \to M$ becomes an R-algebra, the *endomorphism algebra* of the module M, when we choose composing of maps as the multiplication.

Sub- and quotient structures of algebras can be defined in a natural way, such that they preserve both the module structure and the multiplication of the algebra.

DEFINITION 3.2. A submodule B of an R-algebra A is an *R-subalgebra*, if it is a submodule of the module A and also satisfies

$$x \cdot y \in B \qquad \text{for all } x, y \in B.$$

A submodule I is called an *R-ideal*, if

$$a \cdot x \in I \quad \text{ja} \quad x \cdot a \in I \qquad \text{for all } a \in A \text{ and } x \in I.$$

If an algebra A has an ideal I, we can construct the *quotient algebra* A/I similarly as with modules. The multiplication in the quotient algebra satisfies $(a + I) \cdot (b + I) = ab + I$. The fact that this multiplication is well defined, can be proved similarly for rings, since the proof doesn't use the associativity or the neutral element of the multiplication of A.

DEFINITION 3.3. Let A and B be two R-algebras. A linear map $\varphi\colon A \to B$ is called an *R-algebra homomorphism*, if

$$\varphi(x \cdot y) = \varphi(x) \cdot \varphi(y) \qquad \text{for all } x, y \in A.$$

If A and B are unitary, we also demand that $\varphi(1_A) = 1_B$.

The kernel of an algebra homomorphism is an ideal, and we have an analogous homomorphism theorem for algebras, as we do for modules.

3.2. Bases of algebras. If an R-algebra is free as an R-module, it is called a *free algebra*. Thus a free algebra has a basis. It turns out that the multiplication table for the basis elements determines the multiplication of the whole algebra.

PROPOSITION 3.4. *Suppose that A is a free R-algebra, with basis B.*

i) *The algebra A is associative, if and only if $(ab)c = a(bc)$ for all basis elements $a, b, c \in B$.*
ii) *The algebra A has a unit element 1, if and only if $1 \cdot a = a$ and $a \cdot 1 = a$ for all $a \in B$.*
iii) *The algebra A is commutative, if and only if $ab = ba$ for all $a, b \in B$.*

TODISTUS. We prove item (iii); suppose that the basis elements commute with each other. Let $x, y \in A$ be arbitrary. They can be written as linear combinations of basis elements: $x = \sum_i x_i b_i$ and $y = \sum_j y_j b_j$. Using the bilinearity of the algebra multiplication we obtain

$$\begin{aligned} x \cdot y &= \sum_i x_i b_i \cdot \sum_j y_j b_j = \sum_i x_i \left(\sum_j y_j (b_i \cdot b_j) \right) = \sum_{i,j} x_i y_j (b_i \cdot b_j) \\ &= \sum_{i,j} x_i y_j (b_j \cdot b_i) = \sum_j y_j \left(\sum_i x_i (b_j \cdot b_i) \right) = \sum_j y_j b_j \cdot \sum_i x_i b_i = y \cdot x. \end{aligned}$$

Thus the algebra is commutative. Observe that above we used the commutativity of the coefficient ring. The other direction of item (iii) is obvious, and the other claims can be proved analogously. □

PROPOSITION 3.5. *Let A be a free R-algebra, with basis B. Suppose also that C is another R-algebra, and $\varphi\colon A \to C$ is an R-linear map. Then the map φ is an algebra homomorphism, if and only if for all basis elements $a, b \in B$ we have that $\varphi(a \cdot b) = \varphi(a) \cdot \varphi(b)$.*

TODISTUS. Let $x = \sum_i x_i b_i$ and $y = \sum_j y_j b_j$ be arbitrary elements of the algebra A. Since the map φ is linear and the algebra multiplication is bilinear, we have

$$\begin{aligned} \varphi(x \cdot y) &= \varphi \left(\sum_{i,j} x_i y_j (b_i \cdot b_j) \right) = \sum_{i,j} x_i y_j \, \varphi(b_i \cdot b_j) = \sum_{i,j} x_i y_j (\varphi(b_i) \cdot \varphi(b_j)) \\ &= \sum_i x_i \varphi(b_i) \cdot \sum_j y_j \varphi(b_j) = \varphi(x) \cdot \varphi(y). \end{aligned}$$

Thus φ is an algebra homomorphism. The other direction of the claim clearly holds. □

Let A be a free R-algebra, with basis B. Every product $b_i \cdot b_j$ of basis elements can be written as a linear combination of basis elements in the form

$$b_i \cdot b_j = \sum_k c_{ij}^k b_k.$$

The constants $c_{ij}^k \in R$ (here k is an upper index, not an exponent) are called the *structural constants of the algebra with respect to the basis B*. From bilinearity of the multiplication it follows that knowing the structural constants is enough to determine the multiplication of the algebra, since

$$\sum_i x_i b_i \cdot \sum_j y_j b_j = \sum_{i,j} x_i y_j (b_i \cdot b_j) = \sum_{i,j,k} x_i y_j c_{ij}^k b_k. \tag{3}$$

The formula above gives the product of linear combinations of basis elements in a general form. Conversely, the family of structural constants (c_{ij}^k) can be chosen

arbitrarily for every i and j, and the formula (3) then defines a bilinear multiplication. We collect these facts in the following proposition.

PROPOSITION 3.6. *Let M be a free R-module, with basis $B = \{b_i\}_{i\in I}$. Let also $(c_{ij}^k)_{k\in I}$ be a finitely supported family of elements of the ring R for every $i, j \in I$. Then in the module M we can define a unique R-bilinear multiplication, whose structural constants with respect to the basis B are exactly the constants c_{ij}^k.*

EXAMPLE 3.7. Consider the two dimensional real vector space $\mathbb{R}^2$. Denote the elements of the natural basis of this space by $1 = (1, 0)$ and $i = (0, 1)$ and define the multiplication table of the basis vectors as follows:

$\cdot$	1	i
1	1	i
i	i	-1

The $\mathbb{R}$-algebra defined using this multiplication table is clearly unitary, associative and commutative. This defines the *complex number algebra*, which thus is a two dimensional algebra with real coefficients. The structural constants of the complex number algebra are listed in the table below, where we have denoted $c_{xy}^z = c_{xy}(z)$.

(x, y)	$(1, 1)$	$(1, i)$	$(i, 1)$	(i, i)
$c_{xy}(1)$	1	0	0	-1
$c_{xy}(i)$	0	1	1	0

Because the complex number algebra is commutative and every non-zero element has an inverse element, the algebra is a field.

EXAMPLE 3.8. *The quaternions.* In the year 1843 William Hamilton[1], discovered a multiplication table for a four dimensional real algebra $\mathbb{H}$, which he called the quaternions[2]. An important feature of the quaternions is that every non-zero element has an inverse element, and thus division is possible. Hamilton had for a long time worked on finding a certain three dimensional real algebra, when once walking on the streets of Dublin, he suddenly realized that he can make his idea work, if he added another dimension to the space. He was so excited of this discovery that he immediately wrote the multiplication rules of the quaternion basis to the stone paving of the Brougham bridge.

If the basis elements of the quaternion algebra are denoted by the symbols 1, i, j and k, the multiplication table is the following:

$\cdot$	1	i	j	k
1	1	i	j	k
i	i	-1	k	$-j$
j	j	$-k$	-1	i
k	k	j	$-i$	-1

From the multiplication table we see that the quaternion algebra is associative and unitary. Moreover every non-zero element has an inverse element: for example the inverse element of $1 + j$ is $\frac{1}{2}(1 - j)$, since

$$(1 + j) \cdot \frac{1}{2}(1 - j) = \frac{1}{2}(1 - j + j - j^2) = 1.$$

[1]William Rowan Hamilton, 1805–1865, an Irish physicist and mathematician

[2]quaternio = four-tuple (lat.)

However, the quaternion algebra is not commutative, so it is not a field. It is called a *division algebra* (compare division ring). As in the complex number algebra, every basis element (apart from 1) is a square root of -1.

The quaternions were used in describing the geometry of three dimensional space already before the concept of a vector space was defined; using quaternions we can for example describe the dot and cross products and rotations in three dimensional space. Hamilton's original motivation was to invent an algebra, where three dimensional rotations could be described by multiplication, analogously as in the complex plane.

A normed algebra is an algebra, such that in the underlying vector space one can define a norm, which is compatible with the multiplication (For example the usual norm of complex numbers: $|x+yi| = \sqrt{x^2+y^2}$.) One can prove that (up to isomorphism) there exist only four normed real division algebras: the real numbers $\mathbb{R}$, the complex numbers $\mathbb{C}$, the quaternions $\mathbb{H}$ and the *octonions* $\mathbb{O}$, which is an eight dimensional non-associative division algebra. If the basis elements of the octonions are denoted by $\{1, i_0, \ldots, i_6\}$, then $\pm i_k$ is a square root of -1 for every k.

KUVA 3. The multiplication tables of the quaternions and octonions

The multiplication tables of the basis elements of the quaternions and octonions are described in the picture. The product of two elements is the third element, which is found on the same line. The direction of the arrow tells you the sign. For example for the quaternions one has $j \cdot i = -k$, and for the octonions $i_0 \cdot i_1 = i_3$ and $i_1 \cdot i_4 = -i_2$.

3.3. Lisätietoa: ryhmä- ja monoidialgebrat. Olkoon $(G, *)$ jokin ryhmä, ja olkoon R rengas. Tarkastellaan vapaata R-modulia $R^{(G)}$. Tämän modulin luonnollisen kannan muodostavat alkiot e_g, missä $g \in G$, ja kukin näistä voidaan samastaa ryhmän alkion g kanssa. Koska kannan alkiot tällöin kuuluvat ryhmään G, niille voidaan määritellä luonnollinen kertolasku.

DEFINITION 3.9. *Ryhmäalgebra* RG on vapaa R-moduli $R^{(G)}$ varustettuna bilineaarisella kertolaskulla, joka toteuttaa ehdon $g \cdot h = g * h$ kaikilla kannan alkioilla $g, h \in G$.

Kahden ryhmäalgebran mielivaltaisen jäsenen tulo on

$$\sum_i a_i g_i \cdot \sum_j b_j h_j = \sum_{i,j} a_i b_j (g_i * h_j).$$

Ryhmäalgebrat ovat liitännäisiä ja ykkösellisiä, mikä seuraa ryhmäkertolaskun ominaisuuksista ja lauseesta 3.4. Samanlainen konstruktio voidaan tehdä lähtien liikkeelle ryhmän sijaan monoidista, jolloin tuloksena on monoidialgebra.

EXAMPLE 3.10. Ryhmien esitysteoriassa tutkitaan homomorfismeja annetulta ryhmältä G jonkin vektoriavaruuden V kääntyvien lineaarikuvausten ryhmään $\mathrm{GL}(V)$. Tällainen kuvaus määrittelee ryhmän G lineaarisen toiminnan avaruudessa V, ja sitä kutsutaan ryhmän *esitykseksi* avaruudessa V. Esityksistä – kuten toiminnoista yleensäkin – on se hyöty, että niitä tutkimalla saadaan paljon tietoa ryhmän rakenteesta.

Nykyisin on tapana sisällyttää esitysteoria modulien teoriaan käyttämällä hyväksi ryhmäalgebran käsitettä. Olkoon V jokin K-kertoiminen vektoriavaruus, ja olkoon $\varphi\colon G \to \mathrm{GL}(V)$ ryhmähomomorfismi, jolloin $\varphi(g)$ on kääntyvä lineaarikuvaus jokaisella $g \in G$. Koska ryhmäalgebra KG on liitännäinen ja ykkösellinen, sitä voidaan pitää renkaana, joka ei kuitenkaan ole vaihdannainen, ellei G ole vaihdannainen. Avaruuteen V voidaan nyt määritellä renkaan KG kanta-alkioiden vasemmanpuoleinen toiminta kaavalla

$$g.x = \varphi(g)(x).$$

Laajentamalla tämä toiminta lineaarisesti koko renkaan KG vasemmaksi toiminnaksi avaruudesta V tulee vasen KG-moduli. Jokaista esitystä φ vastaa nyt yksikäsitteisesti jokin KG-moduli, ja esitysteorian käsitteet voidaan ilmaista modulikäsitteiden avulla.

3.4. Polynomialgebrat. Polynomit muodostavat renkaita, joissa voidaan määritellä luonnollinen skalaarikertolasku. Tähän asti polynomeja on käsitelty tässä materiaalissa varsin epämuodollisesti. Esitetään tässä yhteydessä eräs tapa konstruoida muodollisesti R-kertoiminen polynomialgebra. Konstruktio toimii samalla esimerkkinä vapaiden modulien käytöstä.

Olkoon R rengas ja $I = \{1, 2, \ldots, n\}$ äärellinen indeksijoukko. Ruvetaan määrittelemään polynomialgebraa $R[X_1, \ldots, X_n]$, joka koostuu R-kertoimisista n:n tuntemattoman polynomeista. Tarkastellaan ensin tulomonoidia $M_n = \mathbb{N}^n$, joka koostuu jonoista $\nu = (\nu_1, \ldots, \nu_n)$, missä $\nu_i \in \mathbb{N}$ jokaisella i. Jonojen yhteenlasku määritellään pisteittäin.

Monoidi M_n sisältää konstruoitavan polynomialgebran monomit. Ryhdytään kirjoittamaan mielivaltainen alkio $\nu = (\nu_1, \ldots, \nu_n) \in M_n$ muodossa

$$X^\nu = X_1^{\nu_1} X_2^{\nu_2} \cdots X_n^{\nu_n}.$$

Jokaisesta jonon ν komponentista ν_i tulee siis tuntemattoman X_i muodollinen eksponentti. Jos jokin ν_i on nolla, voidaan vastaava X_i^0 jättää merkitsemättä tuloon. Tällöin $X^{e_i} = X_i$, missä $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ (ykkönen i:nnellä paikalla). Kahden jonon μ ja ν summa vastaa nyt muodollista vaihdannaista tuloa:

$$\mu + \nu = X^{\mu+\nu} = X_1^{\mu_1+\nu_1} X_2^{\mu_2+\nu_2} \cdots X_n^{\mu_n+\nu_n}.$$

Esimerkiksi $(2, 1, 0) + (0, 1, 1) = X_1^2 X_2 \cdot X_2 X_3 = X_1^2 X_2^2 X_3$.

Tarkastellaan sitten monoidialgebraa RM_n. Se on vapaa moduli $R^{(M_n)}$, joten sen alkioina ovat monoidin M_n alkioiden R-kertoimiset lineaarikombinaatiot

$$\sum_\nu a_\nu X^\nu, \qquad \text{missä } a_\nu \in R \text{ ja } \nu \in M_n.$$

Kanta-alkioiden kertolasku määritellään monoidin M_n laskutoimituksen avulla:

$$X^\mu \cdot X^\nu = X^{\mu+\nu} = X_1^{\nu_1+\mu_1} \cdots X_n^{\nu_n+\mu_n}.$$

Tällöin kahden yleisen alkion tulo on

$$\sum_\nu a_\nu X^\nu \cdot \sum_\mu b_\mu X^\mu = \sum_{\nu,\mu} a_\nu b_\mu X^{\nu+\mu}.$$

Huomaa, että algebraan siirryttäessä monoidin yhteenlasku muuttuu algebran kertolaskuksi ja samalla monoidin nolla-alkiosta $X^0 = (0, \ldots, 0)$ tulee algebran ykkösalkio.

DEFINITION 3.11. Monoidialgebra RM_n on *R-kertoiminen n:n tuntemattoman polynomialgebra*. Se on liitännäinen, vaihdannainen ja ykkösellinen R-algebra, ja sitä merkitään $R[X_1, \ldots, X_n]$. Monoidin M_n alkioita kutsutaan *monomeiksi*.

Polynomialgebra koostuu monomien lineaarikombinaatioista. Tyhjä lineaarikombinaatio on algebran nolla-alkio, ja sitä nimitetään *nollapolynomiksi*. Ykkösalkio on monoidin M_n nolla-alkio $X^0 = (0, \ldots, 0)$. Monomin X^ν *aste* on eksponenttien summa $\sum_i \nu_i$, ja polynomin aste on suurin sen sisältämän monomin aste. Nollapolynomin asteeksi määritellään $-\infty$. Esimerkiksi monomin $X_2^5 X_3$ aste on $5 + 1 = 6$. Polynomin f astetta merkitään $\deg(f)$.

Polynomia, jonka aste on 0 tai $-\infty$, nimitetään *vakiopolynomiksi* tai *vakioksi*. Kuvaus $\eta\colon a \mapsto aX^0$ on bijektio renkaan R ja vakiopolynomien välillä, ja sen avulla kerroinrengas voidaan samastaa vakioiden kanssa. Kuvaus η on myös rengashomomorfismi, mistä seuraa, että renkaan skalaarikertolasku yhtyy vakiopolynomien kertolaskuun. Erityisesti η kuvaa renkaan ykkösalkion algebran ykkösalkioksi.

Polynomialgebralle pätee seuraava universaalisuus.

PROPOSITION 3.12. *Olkoon R rengas ja A jokin liitännäinen, vaihdannainen ja ykkösellinen R-algebra. Olkoon lisäksi $(x_1, \ldots, x_n)$ jono A:n alkioita. Tällöin on olemassa yksikäsitteinen algebrahomomorfismi $\varphi\colon R[X_1, \ldots, X_n] \to A$, jolle pätee $\varphi(X_i) = x_i$ jokaisella i.*

TODISTUS. Koska algebra A on liitännäinen ja ykkösellinen, se on kertolaskunsa suhteen monoidi. Määritellään kuvaus $g\colon M_n \to A$ kaavalla

$$g(X^\nu) = x_1^{\nu_1} \cdots x_n^{\nu_1}.$$

Kuvaus g on monoidihomomorfismi monoidilta M_n algebran A multiplikatiiviselle monoidille, sillä

$$\begin{aligned} g(X^\mu \cdot X^\nu) = g(X^{\mu+\nu}) &= x_1^{\mu_1+\nu_1} \cdots x_n^{\mu_n+\nu_n} \\ &= (x_1^{\mu_1} \cdots x_n^{\mu_n}) \cdot (x_1^{\nu_1} \cdots x_n^{\nu_n}) = g(X^\mu) \cdot g(X^\nu), \end{aligned}$$

ja $g(X^0) = x_1^0 \cdots x_n^0 = 1_A$. (Tässä käytettiin hyväksi A:n vaihdannaisuutta.)

Koska $R[X_1, \ldots, X_n]$ on vapaa moduli, jonka kanta on M_n, vapaan modulin universaalisuudesta seuraa, että on olemassa yksikäsitteinen R-lineaarinen kuvaus $\varphi\colon R[X_1, \ldots, X_n] \to A$, jolle pätee $\varphi(X^\nu) = g(X^\nu)$ kaikilla $X^\nu \in M_n$. Lauseen 3.5 nojalla lineaarikuvaus φ on lisäksi algebrahomomorfismi, sillä kannan alkioilla pätee

$$\varphi(X^\nu \cdot X^\mu) = g(X^\nu \cdot X^\mu) = g(X^\nu) \cdot g(X^\mu) = \varphi(X^\nu) \cdot \varphi(X^\mu).$$

Lisäksi jokaisella i pätee $\varphi(X_i) = g(X^{e_i}) = x_i$.

Olkoon sitten φ' toinen algebrahomomorfismi, joka toteuttaa lauseen oletukset. Koska φ' säilyttää kertolaskun, täytyy päteä

$$\varphi'(X^\nu) = \varphi'(X_1^{\nu_1} \cdots X_n^{\nu_n}) = \varphi'(X_1)^{\nu_1} \cdots \varphi'(X_n)^{\nu_n} = x_1^{\nu_1} \cdots x_n^{\nu_1}.$$

Näin ollen kuvaukset φ' ja φ yhtyvät monoidin M_n alkioilla, joten φ:n yksikäsitteisyydestä seuraa $\varphi' = \varphi$. □

DEFINITION 3.13. Edellisen lauseen kuvausta φ kutsutaan algebran A alkioihin $x_1, \ldots, x_n$ liittyväksi *sijoitushomomorfismiksi*. Polynomin $f = \sum_\nu a_\nu X^\nu$ arvo sijoitushomomorfismissa on

$$\varphi(f) = \sum_\nu a_\nu x_1^{\nu_1} \cdots x_n^{\nu_n}.$$

Tätä arvoa merkitään myös $f(x_1, \ldots, x_n)$.

Sijoitushomomorfismin avulla voidaan määritellä algebran A *polynomifunktio* $(x_1, \ldots, x_n) \mapsto f(x_1, \ldots, x_n)$. Tässä kohdassa on syytä huomata ero polynomin ja sen määräämän polynomifunktion välillä. Olkoon esimerkiksi $f = X^2 + X \in \mathbb{Z}_2[X]$. Nyt f ei ole nollapolynomi, mutta $f(x) = 0$ kaikilla $x \in \mathbb{Z}_2$, eli f:n määräämä funktio algebrassa $\mathbb{Z}_2$ on nollafunktio.

Jos $\varphi\colon R[X] \to A$ on alkioon $\alpha \in A$ liittyvä sijoitushomomorfismi ja $\varphi(f) = 0$, alkiota α nimitetään polynomin f *juureksi*. Polynomifunktion käsitteen avulla ilmaistuna α on polynomin f juuri, jos se on funktion $x \mapsto f(x)$ nollakohta eli $f(\alpha) = 0$.

Tässä luvussa määriteltiin polynomit äärellisen muuttujajoukon $\{X_1, \ldots, X_n\}$ suhteen. On myös mahdollista valita indeksijoukko I äärettömäksi. Tällöin monomimonoidi $M_I = \mathbb{N}^{(I)}$ koostuu alkioperheistä, joissa on vain äärellinen määrä nollasta poikkeavia alkioita. Muuten konstruktio etenee aivan samalla tavalla.

3.5. Lisätietoa: Lien algebrat. Lien algebrat tarjoavat tärkeän esimerkin epäliitännäisistä algebroista. Useimmiten Lien algebrat esiintyvät Lien ryhmien yhteydessä, jotka puolestaan kuvaavat jatkuvien objektien symmetrioita; eräs esimerkki on ympyrän symmetriaryhmä. Lien ryhmillä on paljon käyttöä paitsi puhtaan matematiikan puolella myös teoreettisessa fysiikassa. Lien algebroja käytetään myös Lie-tyypin äärellisten ryhmien määrittelemiseen (ks. lause **??**).

DEFINITION 3.14. Olkoon $\mathfrak{L}$ jokin K-vektoriavaruus. Oletetaan, että avaruudessa $\mathfrak{L}$ on määritelty bilineaarinen tulo $(x, y) \mapsto [xy]$, jolle pätee

(LA1) $[xx] = 0$ kaikilla $x \in \mathfrak{L}$

(LA2) $[x[yz]] + [y[zx]] + [z[xy]] = 0$ kaikilla $x, y, z \in \mathfrak{L}$.

Tällöin avaruutta $\mathfrak{L}$ kutsutaan *Lien algebraksi*.

Ehtoa (LA1) nimitetään *alternoivuudeksi* ja ehtoa (LA2) *Jacobin identiteetiksi*. Lien algebran kertolasku ei yleensä ole liitännäinen eikä vaihdannainen. Sen sijaan ehdon (LA1) ja kertolaskun bilineaarisuuden perusteella pätee

$$0 = [(x+y)(x+y)] = [xx] + [xy] + [yx] + [yy] = [xy] + [yx],$$

mistä seuraa

(LA1′) $[xy] = -[yx]$ kaikilla $x, y \in V$.

Viimeksi mainittu ehto on nimeltään *antisymmetrisyys*. Jos kerroinkunnan karakteristika ei ole 2, ehdot (LA1) ja (LA1′) ovat yhtäpitäviä: tällöin nimittäin ehto (LA1) saadaan asettamalla $x = y$ yhtälössä $[xy] = -[yx]$.

Liitännäisten algebrojen avulla voidaan tuottaa runsaasti esimerkkejä Lien algebroista. Olkoon A jokin liitännäinen K-algebra, esimerkiksi K-kertoimisten neliömatriisien muodostama algebra. *Lien kommutaattori* määritellään kaavalla

$$[x, y] = xy - yx \qquad \text{kaikilla } x, y \in A.$$

Algebrasta A tulee Lien algebra kertolaskun $(x, y) \mapsto [x, y]$ suhteen. Tämä kertolasku on nimittäin selvästi bilineaarinen, ja sille pätee ehto (LA1). Jacobin identiteetin tarkistamiseksi todetaan, että

$$\begin{aligned}[x,[y,z]] &= [x, yz - zy] = (xyz - xzy) - (yzx - zyx)\\ [y,[z,x]] &= [y, zx - xz] = (yzx - yxz) - (zxy - xzy)\\ [z,[x,y]] &= [z, xy - yx] = (zxy - zyx) - (xyz - yxz).\end{aligned}$$

Kun lasketaan yllä olevat lausekkeet yhteen, saadaan tulokseksi 0. Liitännäisyyttä käytettiin siihen, että kolminkertaiset tulot voitiin kirjoittaa ilman sulkeita.

Example 3.15. Palautetaan mieleen, että matriisin x *jälki* on sen diagonaalialkioiden summa: $\operatorname{tr} x = \sum_i x_{ii}$. Tarkastellaan joukkoa

$$\mathfrak{sl}_n(\mathbb{R}) = \{x \in \mathbb{R}^{n\times n} \mid \operatorname{tr} x = 0\}.$$

Koska $\mathfrak{sl}_n(\mathbb{R})$ on lineaarikuvauksen $\operatorname{tr}\colon \mathbb{R}^{n\times n} \to \mathbb{R}$ ydin, se on liitännäisen algebran $\mathbb{R}^{n\times n}$ aliavaruus, joka ei kuitenkaan ole suljettu matriisikertolaskun suhteen (paitsi jos $n = 1$). Jos kuitenkin tarkastellaan avaruutta $\mathbb{R}^{n\times n}$ Lien algebrana kertolaskun $[x, y]$ suhteen, voidaan osoittaa, että $\mathfrak{sl}_n(\mathbb{R})$ on Lien alialgebra. Kaikille matriiseille $x, y \in \mathbb{R}^{n\times n}$ nimittäin pätee

$$\begin{aligned}\operatorname{tr}(xy - yx) &= \sum_{i=1}^{n}\left(\sum_{k=1}^{n} x_{ik}y_{ki} - \sum_{k=1}^{n} y_{ik}x_{ki}\right)\\ &= \sum_{i,k=1}^{n} x_{ik}y_{ki} - \sum_{i,k=1}^{n} x_{ki}y_{ik} = 0.\end{aligned}$$

Lien algebraa $\mathfrak{sl}_n(\mathbb{R})$ nimitetään *(reaalikertoimiseksi) n-ulotteiseksi erityiseksi lineaariseksi algebraksi.*

Jos liitännäinen algebra A on lisäksi vaihdannainen, kommutaattori $[x, y]$ on nolla kaikilla $x, y \in A$. Tämä ominaisuus otetaan Lien algebran vaihdannaisuuden määritelmäksi.

Definition 3.16. Lien algebraa $\mathfrak{L}$ kutsutaan *vaihdannaiseksi*, jos $[xy] = 0$ kaikilla $x, y \in \mathfrak{L}$.

Huomaa, että Lien algebran vaihdannaisuus ei ole aina sama asia kuin yleisen algebran vaihdannaisuus. Jos kerroinkunnan karakteristika ei ole 2, niin Lien algebra $\mathfrak{L}$ on vaihdannainen, jos ja vain jos $[xy] = [yx]$ pätee kaikilla $x, y \in \mathfrak{L}$ eli $\mathfrak{L}$ on vaihdannainen algebra. Kuitenkin karakteristikan ollessa 2 ehto $[xy] = [yx]$ seuraa suoraan ehdosta (LA1′) eikä $\mathfrak{L}$ silti ole välttämättä vaihdannainen Lien algebra.

Proposition 3.17. *Jokainen yksiulotteinen Lien algebra on vaihdannainen.*

Todistus. Oletetaan, että $\mathfrak{L}$ on yksiulotteinen K-kertoiminen Lien algebra. Olkoon v jokin nollasta poikkeava vektori. Avaruus $\mathfrak{L}$ on nyt vektorin v virittämä, joten jokainen alkio on muotoa av, missä $a \in K$. Alternoivuudesta seuraa $[(av)(bv)] = ab[vv] = 0$ kaikilla $a, b \in K$, joten $\mathfrak{L}$ on vaihdannainen. □

Example 3.18. Lien algebrat liittyvät läheisesti *Lien ryhmiin*. Tarkastellaan esimerkkinä Lien ryhmästä jotain reaalikertoimista matriisiryhmää $G \leq \mathrm{GL}_n(\mathbb{R})$.

Tämän ryhmän matriiseja voidaan ajatella avaruuden $\mathbb{R}^{n^2}$ vektoreina, jolloin ryhmässä määritellyille funktioille ja poluille voidaan määritellä raja-arvot ja derivaatat tavalliseen tapaan.

Olkoon $\gamma\colon \mathbb{R} \to G$ derivoituva funktio, jolle pätee $\gamma(0) = 1$ (ykkösmatriisi). Tämä γ on neutraalialkion kautta kulkeva *polku*. Derivaatta $\gamma'(0) \in \mathbb{R}^{n\times n}$ määrää polun *tangenttivektorin* neutraalialkion kohdalla. Voidaan osoittaa, että kaikkien neutraalialkion kautta kulkevien polkujen tangenttivektorit muodostavat avaruuden $\mathbb{R}^{n\times n}$ aliavaruuden $\mathfrak{g}$. Tämä *tangenttiavaruus* on lisäksi suljettu Lien kommutaattorin suhteen, joten se on Lien algebra. Sitä kutsutaan *ryhmän G Lien algebraksi*. Kommutaattorilla on läheinen yhteys konjugointiin ryhmässä G.

KUVA 4. Lien ryhmän G tangenttivektoreita

Jos $x \in \mathfrak{g}$ eli x on jonkin neutraalialkion kautta kulkevan polun γ derivaatta, pätee derivaatan määritelmän mukaan

$$\gamma(t) = 1 + tx + t\epsilon(t),$$

missä $|\varepsilon(t)| \to 0$, kun $t \to 0$. Polulla γ olevia ryhmän alkioita voidaan siis approksimoida Lien algebran alkion tx avulla, kun t on riittävän pieni. Lisäksi voidaan osoittaa, että jos ryhmä G on topologisesti yhtenäinen, mikä tahansa neutraalialkion ympäristö riittää virittämään koko ryhmän. Tämän vuoksi Lien algebraa nimitetään joskus ryhmän "virittäväksi" algebraksi. Toisaalta yllä olevan kaavan alkiota tx voidaan ajatella infinitesimaalisena ryhmän alkiona, ja Lien algebraa nimitetäänkin joskus "infinitesimaaliseksi ryhmäksi", vaikka todellisuudessa Lien algebralla ei ole ryhmän rakennetta.

Esimerkiksi erityinen lineaarinen algebra $\mathfrak{sl}_n(\mathbb{R})$ (ks. esimerkki 3.15) on erityisen lineaarisen ryhmän $\mathrm{SL}_n(\mathbb{R})$ Lien algebra. Ryhmään $\mathrm{SL}_n(\mathbb{R})$ kuuluvat sellaiset $n \times n$ -matriisit, joiden determinantti on 1. Jos $x \in \mathfrak{sl}_2(\mathbb{R})$, niin x on muotoa $\left[\begin{smallmatrix} a & b \\ c & -a \end{smallmatrix}\right]$, ja

$$\det(1 + tx) = \begin{vmatrix} 1 + ta & tb \\ tc & 1 - ta \end{vmatrix} = 1 - t^2(a^2 + bc).$$

Jos parametri t on infinitesimaalisen pieni, toisen potenssin t^2 sisältävä termi voidaan jättää huomiotta. Tällöin huomataan, että matriisin $1 + tx$ determinantti on erittäin lähellä ykköstä, joten kyseinen matriisi approksimoi jotain ryhmän $\mathrm{SL}_2(\mathbb{R})$ alkiota.

KUVA 5. Matriisi $1 + tx$ on lähellä ryhmää $\mathrm{SL}_2(\mathbb{R})$.

Yleisessä tapauksessa Lien ryhmät voivat olla mitä tahansa derivoituvia monistoja. Tällöin tangenttivektorit voidaan määritellä samaan tapaan kuin matriisien tapauksessa. Tangenttivektorien Lien kertolaskua ei kuitenkaan saada suoraan matriisialgebran kommutaattorina, vaan se on johdettava muulla tavalla.

Field extensions

4. Example: construction of a finite field

Before we investigate the theory of field extensions more generally, we look at the structure of finite fields, and one way to construct them. At the same time we familiarize ourselves with some methods we shall need to develop the general theory.

First recall the concepts of the characteristic of a field and a prime subfield (alkukunta) of a field.

DEFINITION 4.1. Let K be a field. The smallest positive integer n, for which it holds that

$$\underbrace{1 + \cdots + 1}_{n \text{ times}} = 0,$$

is called the *characteristic* of K and denoted by $\operatorname{char}(K)$. If such a number doesn't exist, we say that the characteristic is zero.

The characteristic of a field is always a prime number, if it is not zero. If the characteristic p is positive, then the pth multiple of any element is zero, because by the distributive law we have that $(a + \cdots + a) = a(1 + \cdots + 1) = 0$ for all a.

If K is a field then every subfield of K contains all multiples of the unit element. On the other hand if the characteristic $p > 0$, then the multiples of the unit element form a substructure isomorphic with the ring $\mathbb{Z}_p$, which is a field. This field is denoted by $\mathbb{F}_p$ and it is called the *p*rime subfield of the field K. On the other hand if the characteristic of K is 0, then the multiples of the unit element form a structure isomorphic with the ring $\mathbb{Z}$. Every subfield of K contains also the inverse elements of these multiples. Hence a subfield has to contain a field, which is isomorphic with the quotient field $\mathbb{Q}$ of $\mathbb{Z}$. From these observations we obtain the following proposition (see also picture 6).

KUVA 6. Every field contains a unique minimal prime subfield.

PROPOSITION 4.2. *Every field contains a unique minimal subfield, which is called the* prime subfield. *If the characteristic of the field is a prime number p, then this subfield is isomorphic with the field $\mathbb{F}_p$. If the characteristic is zero, then the prime subfield is isomorphic with the field $\mathbb{Q}$ of rational numbers.*

A finite field cannot contain an infinite subfield, thus the characteristic of a finite field is necessarily positive. The characteristic of an infinite field can be positive or zero.

In fact the structure of a finite field is quite accurately determined. The following proposition gives one piece of information in this direction, when we investigate the possible numbers of elements of a finite field.

PROPOSITION 4.3. *If K is a finite field, then $|K| = p^n$, where p is the characteristic of K and n is some positive integer.*

TODISTUS. We identify the prime subfield of K with the field $\mathbb{F}_p$. Now K is an $\mathbb{F}_p$-algebra, with the multiplication of the field as the scalar multiplication. Especially K is a finite $\mathbb{F}_p$-vector space, so it has a finite dimension. Denote the basis of K by $\{b_1, \ldots, b_n\}$. Now every element of K can in a unique way be written in the form

$$x_1 b_1 + x_2 b_2 + \cdots + x_n b_n,$$

where $x_i \in \mathbb{F}_p$ for all i. The number of such linear combinations is the same as the number of possible sequences $(x_1, \ldots, x_n)$ of coefficients, that is, p^n. Thus $|K| = p^n$. □

Later we shall prove that for every such number p^n there exists a field with p^n elements, and that this field is unique up to isomorphism. We conclude this chapter by showing how these kind of fields can be constructed.

We start with the field $\mathbb{F}_p$. Suppose that $f \in \mathbb{F}_p[X]$ is an irreducible polynomial, whose degree is $n > 0$. By Example **??** (which was presented in the lectures previously) the ideal $\langle f \rangle$ generated by f is maximal. Thus the quotient ring $\mathbb{F}_p[X]/\langle f \rangle$ is a field. The elements of this field are the cosets of polynomials $\overline{g} = g + \langle f \rangle$. The zero and unit elements are the cosets of 0 and 1, respectively; these cosets are simply denoted by $\overline{0} = 0$ and $\overline{1} = 1$.

PROPOSITION 4.4. *The number of elements of the field $\mathbb{F}_p[X]/\langle f \rangle$ is p^n.*

TODISTUS. Denote $\mathbb{F}_p[X]/\langle f \rangle = K$. The ideal $\langle f \rangle$ is a submodule of the $\mathbb{F}_p$-module $\mathbb{F}_p[X]$, because $ag \in \langle f \rangle$ for all $g \in \langle f \rangle$ and $a \in \mathbb{F}_p$. Thus K is an $\mathbb{F}_p$-quotient module. By the proof of the previous proposition it is sufficient to show that K has a basis consisting of n elements.

Consider the monomials X^i, where $i \in \{0, \ldots, n-1\}$, and their linear combinations

$$g = \sum_{i=0}^{n-1} a_i X^i, \qquad \text{where } a_i \in \mathbb{F}_p \text{ for all } i.$$

The coset $\overline{g}$ is a linear combination of the set $B = \{1, \overline{X} \ldots, \overline{X}^{n-1}\}$ with the same scalar coefficients a_i. Suppose that $\overline{g} = 0$. Then $g \in \langle f \rangle$, that is, f divides g, but the degree of g is smaller than n, so g has to be the zero polynomial. Thus $a_i = 0$ for all i, from which it follows that the set B is free.

Suppose then that $h \in \mathbb{F}_p[X]$ is an arbitrary element. From the division algorithm for polynomials, it follows that $h = qf + r$, where the degree of r is smaller than n. Now $h - r \in \langle f \rangle$, that is, $\overline{h} = \overline{r}$. On the other hand, r is of the form $a_0 + a_1 X + \cdots + a_{n-1} X^{n-1}$, and thus $\overline{h}$ can be presented as a linear combination of elements of the set B. Hence the set B is a basis for the $\mathbb{F}_p$-vector space K. □

EXAMPLE 4.5. The polynomial $f = X^2 + 1$ is irreducible in the field $\mathbb{F}_3 = \{-1, 0, 1\}$, because none of the elements of the field is a root for the polynomial.

The quotient ring $K = \mathbb{F}_3[X]/\langle f \rangle$ is a field, which consists of linear combinations of the elements 1 and $\overline{X}$:

$$K = \left\{0,\, 1,\, -1,\, \overline{X},\, \overline{X}+1,\, \overline{X}-1,\, -\overline{X},\, -\overline{X}+1,\, -\overline{X}-1\right\}.$$

Thus this is a field with 9 elements, whose binary operations are inherited from the binary operations in the polynomial ring $\mathbb{F}_3[X]$. On the other hand, as in algebras in general, to define multiplication, it is sufficient to determine the multiplication table of the basis elements. We notice that

$$\overline{X}^2 = (\overline{X}^2 + 1) - 1 = 0 - 1 = -1.$$

Thus the element $\overline{X}$ is a square root of the number -1 (more accurately, of the element $-(1 + I)$). Denote now $\overline{X} = i$, then the multiplication in the field K can be thought of as multiplication of complex numbers "modulo 3". For example $(1 + i)^2 = 1 + 2i + i^2 = 2i = -i$.

KUVA 7. A two dimensional extension of the field $\mathbb{F}_3$. The multiplication works similarly as with complex numbers.

5. Tools related to divisibility

In an example in the previous chapter we needed information about irreducibility of a certain polynomial. This is quite usual when dealing with field extensions. In this chapter we discuss conceps related to divisibility in general and prove some criteria concerning irreducibility of polynomials.

5.1. Divisibility in integral domains. Let R be an integral domain[1]. The concepts related to divisibility are defined in R in a similar manner as for integers. An element $a \in R$ *divides* an element $b \in R$, if $b = ac$ for some $c \in R$. Then we denote $a|b$, and we also say that a is a *factor* of b. If $a|b$ and $b|a$, then a and b are *associates* of each other. If two elements are associates, then they have the same factors. Invertible elements (elements which have an inverse with respect to multiplication) are called *units*, and they divide every element of R, because if a is a unit then $b = a(a^{-1}b)$. The proof of the following lemma is left as an exercise.

LEMMA 5.1. *Suppose that $a, b \in R$.*

a) *The elements a and b are associates, if and only if $a = bc$, where c is a unit.*
b) *If $a, b \in R \setminus \{0\}$ are associates and $a = bc$, then c is a unit.*
c) *All units are associates of each other.*

Because every element is divisible by all units and its' associates, these can be regarded as *trivial factors*, which are not taken into account in divisibility considerations. In a general integral domain there exist two types of irreducible elements. Because the zero element is divisible by all elements in any case, it is left outside of this definition.

DEFINITION 5.2. Suppose that $a \in R \setminus \{0\}$ is not a unit. Then a is called *irreducible*, if its' every factor is either a unit or an associate of a.

DEFINITION 5.3. Suppose that $a \in R \setminus \{0\}$ is not a unit. The element a is called *a prime element*, if whenever a divides a product bc, at least one of the elements b and c is divisible by a.

In the ring of integers $\mathbb{Z}$ there are only two units: 1 and -1. An element $n \in \mathbb{Z}$ also has two associates: n and $-n$. Every prime number p is irreducible, because it only has the factors 1, -1, p and $-p$, which all are units or associates of p. The prime numbers and their associates are also the only irreducible integers.

PROPOSITION 5.4. *If $a \in R$ is a prime element, then it is irreducible.*

TODISTUS. Suppose that $a \in R$ is a prime element and $a = bc$ for some $b, c \in R$. Then both b and c divide a. On the other hand a trivially divides the product bc, hence, because a is a prime element, a divides b or c. In the first case a and b are associates, and hence c is a unit. In the second case a and c are associates, and b is a unit. In any case a is divisible only by units and its' own associates. □

[1]Several of these concepts can be defined in general rings, but for simplicity we only discuss integral domains.

The converse result doesn't always hold: an irreducible element is not necessarily a prime element, even though this is true in the case of the integers (by Euclid's lemma). For example in the subring $\mathbb{Z}[i\sqrt{5}] = \{a + bi\sqrt{5} \mid a, b \in \mathbb{Z}\}$ of the complex numbers we have that

$$6 = 2 \cdot 3 = (1 + i\sqrt{5})(1 - i\sqrt{5}).$$

A small calculation shows that the numbers 2, 3 and $1 \pm i\sqrt{5}$ are all irreducible. However, 2 divides the product $(1 + i\sqrt{5})(1 - i\sqrt{5})$, but it doesn't divide any of the factors; hence it is not a prime element.

The greatest common divisor of numbers is also defined in a familiar way.

DEFINITION 5.5. Let $a, b \in R \setminus \{0\}$. An element $d \in R$ is called a *greatest common divisor* of the elements a and b, if the following conditions hold:

i) $d|a$ and $d|b$, that is, d is a common factor of a and b.
ii) If $c|a$ and $c|b$, then $c|d$.

If 1 is a greatest common divisor of elements a and b, we say that a and b are *relatively prime*.

In a general integral domain two elements don't necessarily have a greatest common divisor. Also the greatest common divisor of two elements a and b is not usually unique, and hence the common notation $d = \gcd(a, b)$ in principal cannot be used. From the conditions in the definition it however follows that the greatest common divisors of two elements are associates with each other. The notation $d = \gcd(a, b)$ may then be interpreted to mean that d is some greatest common divisor of a and b, and every other greatest common divisor is obtained by multiplying d with some unit. Especially the notation $\gcd(a, b) = 1$ then means that every greatest common divisor is a unit.

For example in the ring of integers the greatest common divisors of the numbers 30 and 12 are (by definition) the numbers 6 and -6. When talking about positive integers, we usually define the greatest common divisor $\gcd(m, n)$ also to be positive. Then the word "greatest" can also be thought of as meaning the greatest with respect to the usual ordering of the integers.

5.2. Examples of properties related to divisibility. In a field questions concerning divisibility are trivial, because every non-zero element is a unit and hence a factor of every element. On the other hand, in a general integral domain it is not always possible to find a greatest common divisor or write an element as a product of irreducible elements. In the following we present some types of integral domains, with different properties regarding divisibility. The proofs will be omitted, but they can be found in many basic algebra textbooks, for example Nathan Jacobson's *Lectures in Abstract Algebra I. Basic Concepts*.

Unique factorization domains. An integral domain, in which every non-zero element can in a unique way be factored into a product of irreducible elements, is called a *unique factorization domain (UFD)*. The factorization has to be unique with the restriction that the order of the factors doesn't matter and every element may be replaced with an associate element. The ring of integers is a UFD: for example the number 60 has the presentation $2 \cdot 2 \cdot 3 \cdot 5$, which is considered equivalent with the presentation $-5 \cdot 2 \cdot 3 \cdot (-2)$. In a UFD every irreducible element is a prime element. Moreover it is always possible to find a greatest common divisor of two

elements by comparing the factorizations of the elements. It is a bit more difficult to prove that if an integral domain R is a UFD, then also the polynomial ring $R[X]$ is a UFD. From this one can deduce by induction that the ring $R[X_1, \ldots, X_n]$ is a UFD for all n.

Principal ideal domains. In a principal ideal domain (PID) every ideal is generated by one element. From this it follows that there exists a greatest common divisor for any two elements a and b and it can be written in a form $xa + by$. Moreover in a principal ideal domain every strictly increasing chain consisting of principal ideals $\langle a_1 \rangle \subsetneq \langle a_2 \rangle \subsetneq \cdots$ is of finite length. Using this property one can prove that every PID is a UFD. However, the polynomial ring $\mathbb{Z}[X]$ is not a PID, even though it is a UFD.

Euclidean domains. We call an integral domain R a *Euclidean domain*, if we can define a so called *Euclidean function* $\varepsilon\colon R \to \mathbb{N}$. A Euclidean function has to satisfy the following condition:

> If $a, b \in R$ and $b \neq 0$, then there exist $q, r \in R$, such that $a = bq + r$ and either $r = 0$ or $\varepsilon(r) < \varepsilon(b)$.

The purpose of this definition is that in a Euclidean domain it is possible to use the Euclidean algorithm to prove the existence of a greatest common divisor. From this it follows that every Euclidean domain is a PID (compare with the proof of Proposition **??**). For integers a Euclidean function may be defined by the formula $\varepsilon(a) = |a|$, and if K is a field, then in the polynomial ring $K[X]$ the degree of the polynomial gives the value of the Euclidean function. This is also the foundation for the division algorithm for polynomials.

KUVA 8. Different properties concerning divisibility

5.3. Irreducibility of polynomials. We consider some results regarding divisibility of polynomials, to be able to prove some useful criteria for irreducibility. The coefficient ring has to be at least an integral domain, because then the condition $\deg(fg) \geq \deg(f)$ holds for every $g \neq 0$. This property is the foundation for divisibility theory of polynomials.

We begin by proving the division algorithm for polynomials. (This I proved in the lectures earlier, see for example [Hungerford, p. 158].)

PROPOSITION 5.6 (Division algorithm for polynomials). *Suppose that K is a field, and let $f, g \in K[X]$. Suppose that $g \neq 0$. Then there exist unique polynomials $q, r \in K[X]$, for which $f = qg + r$ and* $\deg(r) < \deg(g)$.

TODISTUS. Jakoyhtälö todistetaan samalla tavoin kuin kokonaisluvuilla. Tarkastellaan joukkoa

$$\mathcal{R} = \{f - qg \mid q \in K[X]\}.$$

Tämä joukko on selvästi epätyhjä. Olkoon $r \in \mathcal{R}$ sellainen polynomi, jonka aste on pienin joukossa $\mathcal{R}$. Tällöin $f - qg = r$ jollain $q \in K[X]$. Jos $r = 0$, väite pätee, sillä $\deg(r) = -\infty < \deg(g)$. Muussa tapauksessa merkitään $r = \sum_{i=0}^{n} a_i X^i$ ja $g = \sum_{i=0}^{m} b_i X^i$, missä $a_n \neq 0$ ja $b_m \neq 0$. Jos nyt $\deg(r) \geq \deg(g)$, niin määritellään $q_1 = q + a_n b_m^{-1} X^{n-m}$. Tällöin

$$f - q_1 g = r - a_n b_m^{-1} X^{n-m} g,$$

ja tämän polynomin aste on pienempi kuin $n = \deg(r)$, koska monomin X^n kerroin on 0. Toisaalta $f - q_1 g$ on joukossa $\mathcal{R}$, mikä on ristiriita. Täten $\deg(r) < \deg(g)$.

Yksikäsitteisyyden osoittamiseksi oletetaan, että polynomit q_1, q_2, r_1 ja r_2 toteuttavat lauseen ehdot. Tällöin $q_1 g + r_1 = q_2 g + r_2$, josta edelleen saadaan $(q_1 - q_2)g = r_1 - r_2$. Jos $q_1 \neq q_2$, niin polynomin $(q_1 - q_2)g$ aste on vähintään $\deg(g)$, joka on suurempi kuin $\deg(r_1 - r_2)$. Tämä on mahdotonta, joten $q_1 = q_2$, mistä seuraa, että $r_1 = r_2$. □

Remark. In the proof one only needed the invertibility of the coefficient b_m. The result thus holds in any integral domain K, as soon as the leading coefficient of the polynomial g is a unit.

From the division algorithm it follows that the polynomial ring (with one indeterminate) is a principal ideal domain, when the coefficient ring is a field . Using this fact one can prove that every irreducible polynomial is a "prime polynomial", from which it follows that unique factorization holds in the polynomial ring.

LEMMA 5.7. *Let K be a field $f, g, h \in K[X]$. Suppose that f is irreducible and $f|(gh)$. Then $f|g$ or $f|h$.*

TODISTUS. In a previous example we have shown that $\langle f \rangle$ is a maximal ideal. From this it follows that $\langle f \rangle$ is a prime ideal. The condition $f|(gh)$ means that $gh \in \langle f \rangle$. Then either $g \in \langle f \rangle$ or $h \in \langle f \rangle$, that is, $f|g$ or $f|h$. □

PROPOSITION 5.8. *If K is a field, then the polynomial ring $K[X]$ is a unique factorization domain.*

TODISTUS. The proof is again similar as for integers. Suppose that $f \in K[X] \setminus \{0\}$ is not irreducible or a unit. Then $f = f_1 f_2$ for some $f_1, f_2 \in K[X]$, neither of which is a unit. Because K is a field, it follows that f_1 and f_2 are not constant polynomials, and then that both have degree strictly smaller that $\deg(f)$. If f_1 or f_2 is not irreducible, we continue by finding again non-trivial factors. The process ends eventually, because the degree cannot decrease infinitely many times. Finally we obtain a presentation $f = f_1 f_2 \cdots f_r$, where every f_i is irreducible.

Then suppose that $f = f_1 \cdots f_r = g_1 \cdots g_s$, where every f_i and g_i is irreducible. Now f_1 divides the product $g_1 \cdots g_s$, and because f_1 is irreducible, it follows from the previous lemma that f_1 divides one of the polynomials g_i. By renumbering, we may suppose that $f_1 \mid g_1$. On the other hand, g_1 is irreducible, and thus f_1 and g_1 are associates. From this it follows that $f_2 \cdots f_r = u g_2 \cdots g_s$, where u is a unit. By induction we can deduce that $r = s$ and that f_i and g_i are associates for every i. □

In the proof we used the fact that in the ring $K[X]$ irreducible elements are also prime elements. Every UFD has this property.

LEMMA 5.9. *In a unique factorization domain every irreducible element is a prime element.*

TODISTUS. Suppose that p is an irreducible element, which divides the product ab. We write a and b as products of irreducible elements in the form $a = a_1 a_2 \cdots a_r$ and $b = b_1 b_2 \cdots b_s$. Now $a_1 \cdots a_r b_1 \cdots b_s$ is a factorization of ab into irreducible factors. From the uniqueness of the factorization it follows that p is one of the elements a_i or b_i or an associate of those. (Otherwise there would exist another

factorization, which consists of p and the irreducible factors of the element ab/p.) Thus $p|a$ or $p|b$. □

Next we prove some practical criteria for irreducibility. The first one is familiar to many from high school.

PROPOSITION 5.10 (Criterion for rational roots). *Let R be a UFD, and let K be its' division field. Suppose that the polynomial $f = a_0 + \cdots + a_n X^n \in R[X]$ has a root $p/q \in K$, where $\gcd(p, q) = 1$. Then p divides the coefficient a_0, and q divides the coefficient a_n.*

TODISTUS. We multiply both sides of the equation $f(p/q) = 0$ by the number q^n, and we obtain

$$a_0 q^n + a_1 p q^{n-1} + a_2 p^2 q^{n-2} + \cdots + a_n p^n = 0.$$

By taking p as a common factor and moving the terms we obtain

$$a_0 q^n = -p(a_1 q^{n-1} + a_2 p q^{n-2} + \cdots + a_n p^{n-1}).$$

From the equation above we see that p divides the product $a_0 q^n$. Because R is a UFD, the element p can be presented as a product of irreducible elements, and each one of those divides the product $a_0 q^n$. By Lemma 5.9 every irreducible factor of p is a prime element, but by assumption p doesn't have common non-trivial factors with the element q^n. Thus every irreducible factor of p divides the element a_0, from which it follows that $p|a_0$.

Analogously from the equation

$$q(a_0 q^{n-1} + a_1 p q^{n-2} + \cdots + a_{n-1} p^{n-1}) = -a_n p^n$$

we can deduce that $q|a_n$. □

This criterion can be used as an irreducibility test for polynomials of degree at most three, as the following example shows.

EXAMPLE 5.11. Consider the polynomial $f = 3X^3 + 3X - 1 \in \mathbb{Z}[X]$. This polynomial is not divisible by any integer, so if it is not irreducible, then it is of the form $f = (aX + b)g$, where $g \in \mathbb{Z}[X]$ is a polynomial of degree two. Then it has the rational root $-b/a$. By the above criterion the rational roots of f belong to the set $\{\pm 1, \pm 1/3\}$. However, none of these is a root of f, thus f is irreducible.

To obtain other criteria we need some lemmas. The following observation is often useful, we omit the proof here.

LEMMA 5.12. *Let R be a ring and suppose that I is an ideal of the ring R. Then the map $R[X] \to (R/I)[X]$, where $\sum_i a_i X^i \mapsto \sum_i (a_i + I)X^i$, is a surjective ring homomorphism.*

In what follows, we denote by $\overline{f}$ the image of $f \in R[X]$ in the map of the previous lemma. Thus the polynomial $\overline{f} \in (R/I)[X]$ is obtained by replacing the coefficients of f by their cosets. Typically we choose a prime ideal I, because then the quotient ring is an integral domain.

DEFINITION 5.13. Let R be a UFD, and let $f \in R[X]$. If the greatest common divisor of the coefficients of the polynomial f is 1, we say that f is *primitive.*

This concept is needed to separate such reducible polynomials, which are divisible by some constant (other than a unit). For example the non-primitive polynomial $2X + 2$ factors non-trivially in the ring $\mathbb{Z}[X]$, but it is irreducible in the ring $\mathbb{Q}[X]$, because there the factor 2 is a unit.

LEMMA 5.14. *Suppose that R is a UFD, and let $f, g \in R[X]$. If f and g are primitive, then also fg is primitive.*

TODISTUS. Suppose the contrary that f and g are primitive but fg is not. Then there exists an element $p \in R$, which divides all the coefficients of the product polynomial fg, and p is not a unit. Since R is a UFD, we may assume that p is a prime element. Now the quotient ring $R/\langle p \rangle$ is an integral domain, from which it follows that also $R/\langle p \rangle[X]$ is an integral domain.

Let $\overline{f}, \overline{g} \in R/\langle p \rangle[X]$ be those polynomials which are obtained from f and g by replacing the coefficients with their cosets with respect to the ideal $\langle p \rangle$ (compare Lemma 5.12). Because f and g are primitive, the element p doesn't divide all the coefficients of neither of these polynomials. From this we see that $\overline{f} \neq 0$ and $\overline{g} \neq 0$. Because $R/\langle p \rangle[X]$ is an integral domain, we have that $\overline{fg} = \overline{f} \cdot \overline{g} \neq 0$. But this means that p doesn't divide all the coefficients of the product fg, which is a contradiction. Thus fg is primitive. □

PROPOSITION 5.15 (The Gauss lemma[1]). *Let R be a UFD, with quotient field K. Then $f \in R[X]$ is irreducible, if and only if f is primitive and irreducible in the ring $K[X]$.*

TODISTUS. Suppose first that f is primitive and irreducible in the ring $K[X]$. If f is not irreducible in the ring $R[X]$, then $f = gh$ for some $g, h \in R[X]$, where neither g nor h is a unit in the ring $R[X]$. However, if f is irreducible in the ring $K[X]$, either g or h is a constant. This constant divides all the coefficients of the polynomial $gh = f$, which is impossible since f is primitive. Thus f is irreducible in the ring $R[X]$.

Suppose then that f is irreducible in the ring $R[X]$. It can be written in the form $f = cf_1$, where c is the greatest common divisor of the coefficients of f and f_1 is primitive. From irreducibility it now follows that c is a unit in $R[X]$, and thus also f is primitive.

Then we make the counterassumption that f is not irreducible in the ring $K[X]$. Then $f = gh$ for some $g, h \in K[X]$, where neither g nor h is a constant. By expanding the coefficients of the product gh, such that the denominators are the same, this product may be written in the form $gh = a/b \cdot g_1 h_1$, where $g_1, h_1 \in R[X]$ are primitive and $a, b \in R$ are relatively prime. Then $bf = ag_1h_1$. By the previous lemma, the product g_1h_1 is primitive. Now b is the greatest common divisor of the coefficients of the polynomial bf (because f is primitive), and a is the greatest common divisor of the coefficients of the polynomial ag_1h_1, from which it follows that b and a are associates. Because $\gcd(a, b) = 1$, this is impossible, unless a and b are units in R. In the latter case we can write $f = (ag_1)(b^{-1}h_1)$, but then f is not irreducible in the ring $R[X]$. This is a contradiction, from which it follows that f is irreducible in the ring $K[X]$. □

[1]Sometimes also the Lemma 5.14 is called the Gauss lemma.

PROPOSITION 5.16 (The Eisenstein criterion). *Let R be a UFD, with quotient field K, and let $f = a_0 + \cdots + a_n X^n \in R[X]$. The polynomial f is irreducible in the ring $K[X]$, if one of the following conditions holds:*

a) *Some prime element $p \in R$ divides the coefficients $a_0, \ldots, a_{n-1}$ but not the coefficient a_n, and p^2 doesn't divide the coefficient a_0.*
b) *Some prime element $p \in R$ divides the coefficients $a_1, \ldots, a_n$ but not the coefficient a_0, and p^2 doesn't divide the coefficient a_n.*

TODISTUS. Suppose that condition a) holds. We may assume that f is primitive. (Otherwise we divide f with the greatest common divisor of its' coefficients, which doesn't affect irreducibility with respect to the field K.) Let $\overline{f}$ be the polynomial of the ring $R/\langle p\rangle[X]$, which is obtained from f by replacing the coefficients with their cosets with respect to the ideal $\langle p\rangle$ (see Lemma 5.12). By condition a) we have that $\overline{f} = \overline{a}_n X^n$. Suppose that f is not irreducible in the ring $K[X]$. By the Gauss lemma $f = gh$, where $g, h \in R[X]$. Because f is primitive, neither g nor h is a constant. Now $\overline{g} \cdot \overline{h} = \overline{f} = \overline{a}_n X^n$, from which it follows that both $\overline{g}$ and $\overline{h}$ are of the form cX^i. If neither of the polynomials $\overline{g}$ or $\overline{h}$ is a constant, then p divides the constant terms of the polynomials g and h. However, then p^2 divides a_0, which is against the assumption. Thus we may assume that for example $\overline{g}$ is a constant. However, g is not a constant, so p divides the coefficient of the highest order term of g. Then p also divides the coefficient a_n, which again is against our assumption. Thus the polynomial f is irreducible in the ring $K[X]$. The case b) is proved in a similar manner. □

EXAMPLE 5.17. The polynomial $X^5 - 12X^3 + 2X + 2$ is seen to be irreducible in the ring $\mathbb{Q}[X]$, when we choose $p = 2$ in the Eisenstein criterion. The polynomial $3X^3 + 3X - 1$ of one previous example is also irreducible, which we notice by choosing $p = 3$. The Eisenstein criterion doesn't say anything about the polynomial $X^4 + 2X + 4$, though.

PROPOSITION 5.18. *Let R be a UFD, with quotient field K, and let $f \in R[X]$. Suppose that $p \in R$ is a prime element, which doesn't divide the leading coefficient of f. If $\overline{f}$ is irreducible in the ring $R/\langle p\rangle[X]$, then f is irreducible in the ring $K[X]$.*

TODISTUS. We may again suppose that f is primitive. If f is not irreducible in the ring $K[X]$, then by the Gauss lemma it can be factored also in the ring $R[X]$. If $f = gh$, where $g, h \in R[X]$, then $\overline{f} = \overline{g} \cdot \overline{h}$ by Lemma 5.12. Furthermore neither g nor h is a constant, because f is primitive. If $\overline{f}$ is irreducible, then $\overline{g}$ or $\overline{h}$ is a unit in the integral domain $R/\langle p\rangle[X]$. Units in a polynomial ring are constants, so because g and h are not constants, we have that p divides the leading coefficient of either g or h. This is impossible, since the leading coefficient of f is not divisible by p. Thus f is irreducible in the ring $K[X]$. □

The converse result doesn't hold: for example $X^3 + X + 1 \in \mathbb{Z}[X]$ is irreducible, but in the ring $\mathbb{F}_3[X]$ in can be factored as $(X - 1)(X^2 + X - 1)$.

EXAMPLE 5.19. Consider the polynomial $f = 7X^4 - X^3 + 2X + 3 \in \mathbb{Z}[X]$. By writing the coefficients modulo 2, we obtain the polynomial $\overline{f} = X^4 + X^3 + 1 \in \mathbb{F}_2[X]$. Because the polynomial $\overline{f}$ doesn't have roots, it doesn't have any factors of degree one.

Suppose that $f = (X^2 + aX + b)(X^2 + cX + d)$ for some $a, b, c, d \in \mathbb{F}_2$. By doing the multiplication and comparing the coefficients we see that

$$a + c = 1, \quad ac + b + d = 0, \quad ad + bc = 0 \quad \text{and} \quad bd = 1.$$

From the first condition it follows that exactly one of the numbers a and c is zero. From the last condition we see that $b = d = 1$. However, then $ad + bc = 1$, which is a contradiction. Thus $\overline{f}$ is irreducible, and hence also f is irreducible with respect to the field $\mathbb{Q}$. At the same time we have also proved that any polynomial $\sum_{i=0}^{4} a_i X^i$ of degree four is irreducible with respect to $\mathbb{Q}$, as soon as the coefficients a_0, a_3 and a_4 are odd and the others are even.

6. General extensions

In this chapter we familiarize ourselves with the basic concepts regarding field extensions.

6.1. Field extension and its' degree.

DEFINITION 6.1. An *extension* of a field K is any field L which contains K as a subfield. An extension is denoted by L/K ("L over K").

In Chapter 4 we saw that in this situation L is also a K-algebra, with multiplication of L as the scalar multiplication. In fact, an extension of a field K can be defined also as being any K-algebra, which at the same time is a field. However, this kind of extension contains a subfield isomorphic to K, so in practice we have the above situation.

Because an extension of a field K is a K-vector space, it has a well defined dimension.

DEFINITION 6.2. The *degree* of an extension L/K is the dimension of L thought of as a K-vector space. The degree is denoted by $[L : K]$, and it is a positive integer or infinite.

If $[L : K]$ is finite, the extension is called a *finite extension*, otherwise an *infinite extension*.

Examples of field extensions:

- The field $\mathbb{C}$ of complex numbers is a finite extension of the field $\mathbb{R}$ of real numbers. The pair $\{1, i\}$ forms a basis for $\mathbb{C}$, so $[\mathbb{C} : \mathbb{R}] = 2$.
- The field $\mathbb{R}$ is an infinite extension of $\mathbb{Q}$: for example the set $\{2^{1/n} \mid n \in \mathbb{N}\}$ is free with respect to $\mathbb{Q}$, thus the extension $\mathbb{R}/\mathbb{Q}$ cannot have a finite basis. Also $\mathbb{C}$ is an infinite extension of $\mathbb{Q}$.
- In Chapter 4 we investigated extensions $K/\mathbb{F}_p$, where $K = \mathbb{F}_p[X]/\langle f \rangle$ for some irreducible polynomial f. We noticed that the degree of such an extension is the same as the degree of the polynomial f.
- If K is a field, then the polynomial algebra $K[X]$ is an infinite dimensional K-algebra, which contains K (identified with the constant polynomials). The polynomial algebra is not a field, so it is not an extension of K. However, it is an integral domain, whose quotient field is the set $K(X)$ of so called *rational functions* with coefficients in K. This set consists of fractions f/g, where $f, g \in K[X]$ and $g \neq 0$. Since $K(X)$ is a field, it is an extension of the field K. Also it contains the subspace $K[X]$, and thus $[K(X) : K] = \infty$.

The next proposition is concerned with the degrees of successive extensions.

PROPOSITION 6.3. *Let $K \subset L \subset M$ be a sequence of fields. Then*

$$[M : K] = [M : L] \cdot [L : K].$$

If one of the degrees $[M : L]$ and $[L : K]$ is infinite, then $[M : K]$ is infinite.

TODISTUS. Let $\{a_i\}_{i \in I}$ and $\{b_j\}_{j \in J}$ be some bases for the extensions L/K and M/L, respectively. We prove that the set $B = \{a_i b_j \mid i \in I, j \in J\}$ is a basis for the extension M/K. (When indexing the set B we allow also that $a_i b_j = a_k b_l$ for

different index pairs $(i,j) \neq (k,l)$. From the proof it however follows that this situation cannot arise.)

First we have that every $x \in M$ is a linear combination of the form $\sum_j y_j b_j$, where $y_j \in L$ for all j. On the other hand every y_j is of the form $\sum_i x_{ij} a_i$, where $x_{ij} \in K$ for all i. Thus $x = \sum_{i,j} x_{ij} a_i b_j$, and we see that the set B generates M as a K-vector space.

Then we prove that the set B is free. Suppose that $\sum_{i,j} x_{ij} a_i b_j = 0$, where $x_{ij} \in K$ for all i. The set $\{b_j\}$ is free in the L-vector space M, and $\sum_i x_{ij} a_i \in L$ for all j, thus $\sum_i x_{ij} a_i = 0$ for all j. Moreover the set $\{a_i\}$ is free in the K-vector space L, thus $x_{ij} = 0$ for all i and j. Thus the set B is a basis for the extension M/K. From the fact that B is free, it especially follows that $a_i b_j \neq a_k b_l$, when $i \neq k$ or $j \neq l$. Thus we finally obtain that $[M : K] = |B| = |I| \cdot |J| = [L : K][M : L]$. This also includes the case that $[L : K]$ or $[M : L]$ is infinite. □

If $K \subset L \subset M$ is a sequence of fields, the extension L/K is called a *subextension* of the extension M/K. From the previous proposition it follows that if $[M : K] = n$, then the degrees $[M : L]$ and $[L : K]$ are factors of the number n. Especially, if n is a prime number, then the extension M/K doesn't have any non-trivial subextensions L/K.

6.2. Generating. When investigating field extensions, it is often good to know that the extension is generated by certain elements. Generating in this context doesn't mean the same as generating as a vector space. For example, the degree of a finitely generated field extension doesn't have to be finite.

DEFINITION 6.4. Let L be an extension of a field K, and A a set of elements of L.

a) *The subring $K[A]$ generated by A of an extension L/K* is the smallest subring of L, which contains the field K and the subset A.
b) *The subextension $K(A)$ generated by A of an extension L/K* is the smallest subfield of L, which contains the field K and the subset A.

In the case that the set $A = \{a_1, \ldots, a_n\}$ is finite, we simply denote $K[A] = K[a_1, \ldots, a_n]$ and $K(A) = K(a_1, \ldots, a_n)$. Then the field $K(a_1, \ldots, a_n)$ is called a *finitely generated* extension of K.

KUVA 9. The subring $K[A]$ and subextension $K(A)$ of an extension K/L generated by a subset A

Since an arbitrary intersection of subrings is a subring and the same holds for fields, the sets $K[A]$ and $K(A)$ can be defined to be the intersections of such subrings/subfields which contain the field K and the set A. In this way we can prove the existence of the sets $K[A]$ and $K(A)$, which doesn't immediately follow from the definitions.

Polynomial algebras are free finitely generated algebras. From the evaluation homomorphism we obtain a crucial connection between polynomial algebras with coefficients in K and finitely generated field extensions of K. The following proposition makes this more concrete.

PROPOSITION 6.5. *Let L be an extension of a field K, and let $a_1, \ldots, a_n \in L$. Then*

$$K[a_1, \ldots, a_n] = \{f(a_1, \ldots, a_n) \mid f \in K[X_1, \ldots, X_n]\}$$

and

$$K(a_1, \ldots, a_n) = \left\{ \frac{f(a_1, \ldots, a_n)}{g(a_1, \ldots, a_n)} \mid f, g \in K[X_1, \ldots, X_n],\ g(a_1, \ldots, a_n) \neq 0 \right\}.$$

Moreover $K(a_1, \ldots, a_n)$ is the quotient field of the ring $K[a_1, \ldots, a_n]$.

TODISTUS. Let $\varphi\colon K[X_1, \ldots, X_n] \to L$ be the evaluation homomorphism connected with the elements $a_1, \ldots, a_n$. The image of this algebra homomorphism is

$$\operatorname{Im} f = \{f(a_1, \ldots, a_n) \mid f \in K[X_1, \ldots, X_n]\},$$

and it is a subalgebra of the algebra L, thus it is a subring. Every subring M of L, which contains the field K and the elements $a_1, \ldots, a_n$, also contains all K-linear combinations of products formed of these elements, and thus $f(a_1, \ldots, a_n) \in M$ for all $f \in K[X_1, \ldots, X_n]$. Hence $\operatorname{Im}\varphi \subset M$, from which it follows that $K[a_1, \ldots, a_n] = \operatorname{Im}\varphi$. The quotient field Q of the ring $K[a_1, \ldots, a_n]$ consists of elements α/β, where $\alpha, \beta \in K[a_1, \ldots, a_n]$ and $\beta \neq 0$. Every subfield of L, which contains the elements a_i, also contains the ring $K[a_1, \ldots, a_n]$ and also the above mentioned fractions α/β. Thus $K(a_1, \ldots, a_n) = Q$. □

EXAMPLE 6.6. The subextension $\mathbb{Q}(i)$ of the extension $\mathbb{C}/\mathbb{Q}$ consists of fractions $f(i)/g(i)$, where $f, g \in \mathbb{Q}[X]$ and $g(i) \neq 0$. Because $i^2 = -1$, we can restrict to polynomials of first degree. Then

$$\mathbb{Q}(i) = \left\{ \frac{a + bi}{c + di} \mid a, b, c, d \in \mathbb{Q},\ c \neq 0 \text{ or } d \neq 0 \right\}.$$

Furthermore $(c + di)^{-1} = q(c - di)$, where $q = (c^2 + d^2)^{-1} \in \mathbb{Q}$, so we can write

$$\mathbb{Q}(i) = \{a + bi \mid a, b \in \mathbb{Q}\} = \mathbb{Q}[i].$$

The set $\{1, i\}$ generates the $\mathbb{Q}$-vector space $\mathbb{Q}[i]$. Furthermore 1 and i are linearly independent with respect to $\mathbb{Q}$, so $\{1, i\}$ is a basis for $\mathbb{Q}[i]$. Hence the degree of the extension $\mathbb{Q}(i)/\mathbb{Q}$ is $[\mathbb{Q}(i) : \mathbb{Q}] = 2$.

On the other hand, it is possible to show that the extension $\mathbb{Q}(\pi) \subset \mathbb{R}$ of $\mathbb{Q}$ doesn't have a finite basis. Thus finitely generated and finite are different properties.

Using the following proposition, it is often possible to reduce properties of infinitely generated extensions to the case of finitely generated extensions.

PROPOSITION 6.7. *Let L be an extension of a field K, and let $A \subset L$. If $\alpha \in K(A)$, then $\alpha \in K(a_1, \ldots, a_n)$ for some $a_1, \ldots, a_n \in A$. Hence*

$$K(A) = \bigcup \{K(a_1, \ldots, a_n) \mid n \in \mathbb{N},\ a_1, \ldots, a_n \in A\}.$$

TODISTUS. Denote $F = \bigcup \{K(a_1, \ldots, a_n) \mid a_i \in A\}$. Every finitely generated field $K(a_1, \ldots, a_n)$, where $a_i \in A$ for all i, is contained in the field $K(A)$. Hence $F \subset K(A)$. On the other hand, F contains the field K and the set A, so if it is a field, then it has to contain also $K(A)$. Thus it is left to prove that F is a field. Let $\alpha, \beta \in F$. Then $\alpha \in K(a_1, \ldots, a_n)$ and $\beta \in K(b_1, \ldots, b_m)$ for some $a_i, b_i \in A$. Now the elements $\alpha \pm \beta$, $\alpha\beta$ and α/β are in the field $K(a_1, \ldots, a_n, b_1, \ldots, b_m)$, and this field is contained in the union F. Thus F is a field, and $K(A) = F$. □

7. Algebraic extensions

In the old days the purpose of research in algebra was to learn how to solve polynomial equations. Thus a particularly important part of the classical theory of field extensions consists of extensions, whose elements are roots of polynomials (with coefficients in the smaller field). For example every complex number is a root of a polynomial with real coefficients and degree at most two.

7.1. Algebraic extensions and minimal polynomials.

DEFINITION 7.1. Let L be an extension of a field K. An element $\alpha \in L$ is called *algebraic* with respect to K, if there exists a non-zero polynomial $f \in K[X]$, for which $f(\alpha) = 0$. If such a polynomial doesn't exist, then we say that α is *transcendental* with respect to K. If all elements of L are algebraic with respect to K, we say that L is algebraic with respect to K, and the extension L/K is called an *algebraic extension*.

Suppose that L is an extension of a field K and $\alpha \in L$. If α is transcendental with respect to K, it means that for all non-zero polynomials $f \in K[X]$ we have that $f(\alpha) \neq 0$. This means that the kernel of the evaluation map associated with the element α

$$\operatorname{Ker} \varphi_\alpha = \{f \in K[X] \mid f(\alpha) = 0\}$$

is the zero ideal.

Also we see that α is algebraic, if and only if the kernel of the evaluation homomorphism is non-trivial. Because $K[X]$ is a principal ideal domain, the ideal $\operatorname{Ker} \varphi_\alpha$ is generated by some polynomial p, that is, $\operatorname{Ker} \varphi_\alpha = \langle p \rangle$. Because every polynomial in the ideal $\langle p \rangle$ is divisible by p, we see that the degree of p is minimal among the non-zero polynomials of the set $\operatorname{Ker} \varphi_\alpha$. Also we have that all polynomials in the set $\operatorname{Ker} \varphi_\alpha$ which have the same degree as p, are divisible by p, so they differ from p by just a constant factor (thus they are associates of p), and every such polynomial generates the same ideal as p. To define the polynomial uniquely, we can for example demand that the leading coefficient is 1, that is, p is a *monic polynomial*. This polynomial is called the *minimal polynomial* of the element α.

KUVA 10. The kernel of the evaluation polynomial is generated by any non-zero polynomial having minimal degree.

DEFINITION 7.2. Suppose that $\alpha \in L$ is algebraic with respect to K. The *minimal polynomial* of the element α with respect to K is such a non-zero monic polynomial $p \in K[X]$, for which $p(\alpha) = 0$ and whose degree is smallest possible. The minimal polynomial of α with respect to K is denoted by $p = \min(K, \alpha)$.

Remark. Because by definition $p(\alpha) = 0$, we have that $p \in \operatorname{Ker} \varphi_\alpha$. By the reasoning preceding the definition, the minimal polynomial of an element α can be characterized as being the *monic polynomial, which generates the kernel of the evaluation homomorphism associated with the element* α.

EXAMPLE 7.3. The number $\sqrt{2}$ is algebraic with respect to the field $\mathbb{Q}$, because it is a root of the polynomial X^2-2. Because $\sqrt{2}$ is not a rational number, it isn't a root of any polynomial of degree one. Hence $\min(\mathbb{Q},\sqrt{2}) = X^2-2$. Notice that $\min(\mathbb{R},\sqrt{2}) = X-\sqrt{2}$.

One useful fact about the minimal polynomial of an element α is that it tells us the degree of the extension $K(\alpha)$. In the following proposition this fact is collected among other useful properties.

PROPOSITION 7.4. *Let L be an extension of a field K, and suppose that $\alpha \in L$ is algebraic with respect to K. Then*

i) *The minimal polynomial $\min(K,\alpha)$ is irreducible in the ring $K[X]$.*
ii) *If $f \in K[X]$, then $f(\alpha)=0$, if and only if $\min(K,\alpha)$ divides f.*
iii) *$K[\alpha]$ is a field, and $K[\alpha] = K(\alpha)$.*
iv) *If n is the degree of the polynomial $\min(K,\alpha)$, then the elements $1,\alpha,\ldots,\alpha^{n-1}$ form a basis for the extension $K(\alpha)/K$. Especially $[K(\alpha):K] = n < \infty$.*

TODISTUS. Denote $\min(K,\alpha) = p$. We begin with item (ii). If $f(\alpha)=0$ for some $f \in K[X]$, then $f \in \operatorname{Ker}\varphi_\alpha$. Because p generates the ideal $\operatorname{Ker}\varphi_\alpha$, the polynomial f is divisible by p. On the other hand, if $f = pg$ for some $g \in K[X]$, then $f(\alpha) = p(\alpha)g(\alpha) = 0$.

i) Suppose that $p = fg$ for some $f,g \in K[X]$, then we have

$$f(\alpha)g(\alpha) = p(\alpha) = 0.$$

The elements $f(\alpha)$ and $g(\alpha)$ are in the field L. Because L is an integral domain, we have $f(\alpha)=0$ or $g(\alpha)=0$. From item (ii) it follows that f or g is divisible by p. On the other hand, both f and g divide p, so one of them is an associate of p and the other one is a unit (Lemma 5.1). Thus p is irreducible.

iii) By Proposition 6.5 we have $K[\alpha] = \operatorname{Im}\varphi_\alpha$, and on the other hand $\langle p\rangle = \operatorname{Ker}\varphi_\alpha$. From the homomorphism theorem for algebras it now follows that $K[X]/\langle p\rangle \cong K[\alpha]$. Because $K[\alpha] \subset L$ is an integral domain, $\langle p\rangle$ is a prime ideal. On the other hand $K[X]$ is a principal ideal domain, so every non-zero prime ideal is maximal (see for example **??**). From this it follows that actually $K[\alpha]$ is a field. Moreover $K[\alpha] = K(\alpha)$, because $K[\alpha] \subset K(\alpha)$ and $K(\alpha)$ is the smallest field containing K and the element α.

iv) Let $x \in K(\alpha)$. By item (iii) we have that $x = f(\alpha)$ for some $f \in K[X]$. From the division algorithm we see that $f = qp + r$, where $\deg(r) < \deg(p) = n$. Now $f(\alpha) = r(\alpha)$, because $p(\alpha) = 0$. The element $x = r(\alpha)$ can thus be written as a linear combination of the elements $1,\alpha,\ldots,\alpha^{n-1}$. Suppose then that $\sum_{i=0}^{n-1} a_i\alpha^i = 0$ for some $a_i \in K$. Then for the polynomial $g = \sum_{i=0}^{n-1} a_iX^i$ we have that $g(\alpha)=0$, thus by item (ii) we have that p divides g. However, the degree of g is smaller than n, thus g has to be the zero polynomial. This means that $a_i = 0$ for all i and the set $\{1,\alpha,\ldots,\alpha^{n-1}\}$ is free. Hence this set forms a basis for the extension $K(\alpha)$ with respect to the coefficient field K. □

EXAMPLE 7.5. Consider the extension $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$. For the polynomial $f = X^3-2$ we have that $f(\sqrt[3]{2}) = 0$, so the minimal polynomial of the number $\sqrt[3]{2}$ divides f. On the other hand, f is irreducible by the Eisenstein criterion 5.16, so it is the minimal polynomial of the number $\sqrt[3]{2}$. Thus the degree of the extension $\mathbb{Q}(\sqrt[3]{2})$ is 3. Moreover $\mathbb{Q}[\sqrt[3]{2}] = \mathbb{Q}(\sqrt[3]{2})$, thus every element in the extension is

of the form $a + b\sqrt[3]{2} + c\sqrt[3]{4}$. This also holds for the inverse elements x^{-1}, where $x \in \mathbb{Q}[\sqrt[3]{2}]$.

EXAMPLE 7.6. The complex number $\omega = e^{2\pi i/3} = -1/2 + i\sqrt{3}/2$ is a root of the polynomial $X^3 - 1$. This polynomial, however, is not the minimal polynomial of ω, because it can be factored as follows: $X^3 - 1 = (X - 1)(X^2 + X + 1)$. From these factors the latter is irreducible by the criterion for rational roots 5.10, and it has ω as a root. Thus the minimal polynomial of the element ω is $X^2 + X + 1$, and $[\mathbb{Q}(\omega) : \mathbb{Q}] = 2$.

It is not difficult to see that a finite extension is always finitely generated. Previously we noticed that the converse doesn't hold: a finitely generated extension is not necessarily finite. However, the concepts are equivalent, if the extension is algebraic. These thoughts are collected in the following two propositions.

PROPOSITION 7.7. *Let L be a finite extension of a field K. Then L is finitely generated and algebraic with respect to K.*

TODISTUS. Exercise. □

PROPOSITION 7.8. *Let L be an extension of a field K. Suppose that $\alpha_i \in L$ is algebraic with respect to K for all i. Then $K[\alpha_1, \ldots, \alpha_n]$ is a finite extension of K, and for the degree we have that*

$$[K[\alpha_1, \ldots, \alpha_n] : K] \leq \prod_{i=1}^{n} [K(\alpha_i) : K].$$

TODISTUS. We use induction with respect to n. The case $n = 1$ follows from Proposition 7.4. Suppose that the claim holds for the ring $K_1 = K[\alpha_1, \ldots, \alpha_{n-1}]$. (Notice that $K[\alpha_1, \ldots, \alpha_n] = K_1[\alpha_n]$.) Then especially K_1 is a field. Because α_n is algebraic with respect to K and thus also with respect to K_1, it follows from Proposition 7.4 that $K_1[\alpha_n] = K_1(\alpha_n)$ is a field. Moreover by the same proposition $\min(K_1, \alpha_n)$ divides the polynomial $\min(K, \alpha_n)$, hence

$$[K_1(\alpha_n) : K_1] \leq [K(\alpha_n) : K].$$

By the inductive assumption and Proposition 6.3 we have

$$[K_1[\alpha_n] : K] = [K_1[\alpha_n] : K_1] \cdot [K_1 : K] \leq \prod_{i=1}^{n} [K(\alpha_i) : K].$$

Moreover the product on the right hand side is finite by Proposition 7.4. □

From the previous propositions we directly get a condition for an element to be algebraic.

COROLLARY 7.9. *Suppose that L is an extension of a field K. Then $\alpha \in L$ is algebraic with respect to K, if and only if $[K(\alpha) : K]$ is finite. Moreover, L is algebraic, if $[L : K]$ is finite.*

The converse of the latter claim of the corollary doesn't hold. For example the set $\{2^{1/n} \mid n \in \mathbb{N}\}$ generates an algebraic extension of $\mathbb{Q}$, whose degree is infinite.

Now we can prove that being an algebraic extension is a transitive property.

PROPOSITION 7.10. *Let $K \subset L \subset M$ be a sequence of fields. If L/K and M/L are algebraic extensions, then M/K is algebraic.*

TODISTUS. Suppose that $m \in M$. Let $p = a_0 + a_1X + \cdots + a_nX^n$ be the minimal polynomial of the element m with respect to L. Denote $K_1 = K(a_0, \ldots, a_n)$. Because L is algebraic with respect to K and $a_i \in L$ for every i, the extension K_1 is finite by Proposition 7.8. Now $p \in K_1[X]$, thus m is algebraic with respect to K_1. Thus $[K_1(m) : K_1]$ is finite, and

$$[K_1(m) : K] = [K_1(m) : K_1] \cdot [K_1 : K] < \infty.$$

Furthermore, $K(m) \subset K_1(m)$, and thus $[K(m) : K] < \infty$. From Proposition 7.7 it follows that $K(m)$ is algebraic with respect to K. Especially m is algebraic with respect to K, and because m was arbitrary, the whole extension M/K is algebraic. □

7.2. Application: constructions with ruler and compass. The theory presented above can be used in investigating certain classical geometric constructions. These constructions, from which perhaps the squaring of a circle is the most famous, have puzzled mathematicians from the antiquity until the 19th century, when it was proved using algebraic methods that those contructions are not possible.

In ancient Greece geometry had an especially important role in mathematical literature. Because of lack of algebraic notation, geometry was used in the presentation of all mathematical results (which were mainly concerning geometry and number theory). Numbers were represented with lines of different lengths: addition was interpreted as putting two lines one after the other, and multiplying two numbers meant constructing a rectangle, whose sides had those numbers as lengths. In this way one could for example present the distributive law $a(b + c) = ab + ac$ by dividing a rectangle with sides a and $b + c$, into two rectangles which correspond to the products ab and ac.

KUVA 11. A geometric construction representing the distributive law

By a traditional story the philosopher Plato[1] demanded that geometric constructions should be carried out by using only a ruler and a compass. With a ruler you could draw an endless line which goes through two known points, and with a compass you could draw a circle, whose center and radius were known (originally the rules were even more strict, but they are equivalent with the ones presented here). Soon the following three questions emerged, which the Greeks couldn't solve even after countless efforts:

1. *Squaring the circle.* One should produce the side of a square, which has the same area with a given circle.
2. *Doubling the cube.* One should produce the side of a cube, whose volume is twice the volume of a given cube.
3. *Trisecting the angle.* One should produce an angle, which is one third of a given angle.

That the Greeks failed in solving these problems was not due to their incompetence. Namely, in 1837 Pierre Wantzel proved that it is not possible to perform the

[1]Plato (428/427–348/347 BC), an Athenian philosopher, founder of the Academy. At his time, Plato was also a prominent figure in the field of mathematics, although it is not known whether he produced original mathematical results himself.

2nd and 3rd constructions by using only ruler and compass: also the 1st construction is impossible, but this could be proven only after Ferdinand von Lindemann in 1882 proved that the number π is transcendental.

Next we investigate, how these problems concerning geometric constructions can be formulated in the language of algebra. We consider certain subsets $G \subset \mathbb{R}^2$, which here are called *figures*. A *line of a figure* G is a line, which goes through two points of G. A *circle of a figure* G is a circle, whose center is a point in the set G and whose radius equals the distance between two points in the set G.

Suppose we are given a figure $G_0 \subset \mathbb{R}^2$. A *geometric construction starting from the set* G_0 is a finite sequence of figures

$$G_0 \subset G_1 \subset \cdots \subset G_n,$$

where $G_{i+1} = G_i \cup \{P_{i+1}\}$ for every $i < n$ and P_{i+1} is the intersection of some lines or circles of the figure G_i. We say that a figure G *can be constructed from a figure* G_0, if there exists a geometric construction $G_0 \subset \cdots \subset G_n$, where $G_n = G$. The *field of a figure* G, denoted by K_G, is the extension $\mathbb{Q}(A)$, where A contains all x- and y-coordinates of points of G.

The next proposition gives an algebraic condition for constructibility of a figure.

PROPOSITION 7.11. *If a figure G can be constructed from a figure G_0, then*

$$[K_G : K_{G_0}] = 2^n$$

for some $n \in \mathbb{N}$.

TODISTUS. Let $G_0 \subset \cdots \subset G_n = G$ be a geometric construction. From basic analytic geometry we know that each line or circle of a figure G_i can be represented by a polynomial equation, whose coefficients are in the field K_{G_i} and which is of degree at most two. Moreover we know that in order to find the intersection of these lines and circles we have to solve a pair of equations of degree at most two, and the solutions are of the form $x = a_1 + b_1\sqrt{c}$ and $y = a_2 + b_2\sqrt{c}$, where $a_1, a_2, b_1, b_2, c \in K_{G_i}$. Thus $K_{G_{i+1}} \subset K_{G_i}(\sqrt{c})$. Because the minimal polynomial of the number $\sqrt{c}$ with respect to the field K_{G_i} is of degree at most two, we obtain that $[K_{G_{i+1}} : K_{G_i}] \leq 2$. The claim follows from this by induction using Proposition 6.3. □

The converse of the proposition above also holds: If the degree $[K_G : K_{G_0}]$ is a power of two, then the figure G can be constructed from the figure G_0. This fact we don't need when we prove that some construction is impossible, as we do in the following example.

EXAMPLE 7.12. *Trisection of an angle.* We prove that for example the angle of $60°$ cannot be trisected by using only a ruler and compass. We choose the coordinates in such a way that the angle of 60 degrees lies between the line segments OA and OB, where $O = (0, 0)$, $A = (1, 0)$ and $B = (1/2, \sqrt{3}/2)$. Let $G_0 = \{O, A, B\}$, then $K_{G_0} = \mathbb{Q}(\sqrt{3})$ and $[K_{G_0} : \mathbb{Q}] = 2$.

Assume the contrary that the angle AOB can be trisected. Then the intersection point of the line (corresponding to the angle we obtained) and the circle with center in the origin and radius one (which also can be constructed) is (α, β), where $\alpha = \cos 20°$ and $\beta = \sin 20°$. By assumption we can construct a figure G, which contains the point (α, β).

KUVA 12. Trisection of an angle

Next we investigate the coordinate α. From the trigonometric equations we see that

$$\cos(3 \cdot 20^\circ) = 4\cos^3 20^\circ - 3\cos 20^\circ = 4\alpha^3 - 3\alpha.$$

Because $\cos 60^\circ = 1/2$, it follows from this that α is a root of the polynomial $8X^3 - 6X - 1$. Because this polynomial is irreducible with respect to $\mathbb{Q}$ (for example by using the criterion for rational roots), it is an associate of the minimal polynomial $\min(\mathbb{Q}, \alpha)$. Thus the degree of this minimal polynomial is 3, and moreover $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$. By the Propositions 7.11 and 6.3 we have that

$$[K_G : \mathbb{Q}] = [K_G : K_{G_0}] \cdot [K_{G_0} : \mathbb{Q}] = 2^n \cdot 2 = 2^{n+1}$$

for some $n \in \mathbb{N}$, but on the other hand

$$[K_G : \mathbb{Q}] = [K_G : \mathbb{Q}(\alpha)] \cdot [\mathbb{Q}(\alpha) : \mathbb{Q}] = [K_G : \mathbb{Q}(\alpha)] \cdot 3.$$

This clearly is impossible, and thus the figure G cannot be constructed.

This example shows that there cannot exist a general procedure for trisecting an angle by using only ruler and compass. However, some angles can be trisected: for example one can prove that the angle of 30 degrees can be constructed, which means that the trisection of the straight angle is possible.

7.3. Lisätietoa: transkendenttiluvut. Luvun todistamiseksi algebralliseksi riittää löytää polynomi, jonka juuri kyseinen luku on. Transkendenttisuuden todistaminen on sen sijaan työläämpää. Jotkin tapaukset ovat kuitenkin selkeitä: Oletetaan esimerkiksi, että K on kunta, ja tarkastellaan polynomialgebran $K[X]$ jakokuntaa $K(X)$. Tämä kunta on K:n laajennos, ja alkio $X \in K(X)$ on selvästi transkendenttinen K:n suhteen. Kaikkien K-kertoimisten polynomien joukko $K[X]$ nimittäin sisältyy kuntaan $K(X)$, ja alkioon X liittyvä sijoitushomomorfismi $K[X] \to K(X)$ on inkluusiokuvaus. Toisin sanoen: sijoitettaessa alkio X polynomiin f tuloksena on f. Siispä $f(X) = 0$, jos ja vain jos $f = 0$.

Useimmiten transkendenttisista luvuista puhuttaessa tarkoitetaan reaali- tai kompleksilukuja, jotka ovat transkendenttisia rationaalilukujen kunnan suhteen. Nykyään on tunnettua, että transkendenttisia lukuja on olemassa, vieläpä runsain mitoin. On nimittäin varsin helppo osoittaa, että $\mathbb{Q}$-kertoimisia polynomeja on vain numeroituva määrä, jolloin myös niiden juuria on numeroituvan monta (ks. lemma **??**). Siispä *algebrallisten lukujen joukko*

$$\mathbb{A} = \{\alpha \in \mathbb{C} \mid \alpha \text{ on algebrallinen } \mathbb{Q}\text{:n suhteen}\}$$

on numeroituva. Toisaalta kompleksilukujen joukko on ylinumeroituva, joten valtaosa kompleksiluvuista (tai yhtä hyvin reaaliluvuista) on transkendenttisia.

Yllä esitetty päättely on mahdollista tehdä vain, jos kompleksilukujen joukon ylinumeroituvuus tunnetaan. Viimeksi mainitun seikan todisti Georg Cantor vuonna 1878. Kuitenkin jo aiemmin – tarkemmin sanottuna vuonna 1844 – Joseph Liouville oli osoittanut, että eräät hänen löytämänsä luvut ovat transkendenttisia reaalilukuja. Näitä lukuja kutsutaan nykyään Liouvillen luvuiksi. Liouvillen löytö oli ensimmäinen osoitus transkendenttisten lukujen olemassaolosta. Tunnetumpi esimerkki transkendenttisesta luvusta saatiin vuonna 1873, kun Charles Hermite osoitti Neperin luvun e transkendenttisuuden. Hieman myöhemmin, eli

vuonna 1882, Ferdinand von Lindemann onnistui Hermiten menetelmää mukaillen osoittamaan, että myös luku π on transkendenttinen rationaalilukujen suhteen. Lindemannin ja Hermiten todistukset käyttävät analyyttisiä menetelmiä.

Avoimeksi ongelmaksi sen sijaan on jäänyt muun muassa se, onko e algebrallinen vai transkendenttinen laajennoksen $\mathbb{Q}(\pi)$ suhteen (tai yhtä hyvin π laajennoksen $\mathbb{Q}(e)$ suhteen) eli onko olemassa polynomia, jonka kertoimissa saa hyödyntää piin potensseja ja jolla on juurena e.