MATHEMATICAL SCIENCES

A COURSE OF ALGEBRA (PART II)

Frasser C.
Ph.D. in Engineering Sciences, Odessa National Polytechnic University, Ukraine
https://doi.org/10.5281/zenodo.18840568

Abstract

The exploration of algebraic operations finds a natural impetus in the challenge of solving polynomial equations of degree $n \geq 5$. Galois theory provides the conceptual framework for this inquiry, revealing that certain quintic equations with rational coefficients admit no general solution in radicals. This fundamental result, both accessible in statement and deep in implication, justifies a systematic development of the algebraic tools required for its understanding. Part II develops the language and techniques – particularly those concerning polynomials and matrices – necessary to formulate and approach such problems rigorously. It is worthy to mention that this final part is presented from a history-of-science perspective and explores modern abstract algebra with clarity and simplicity, while retaining an advanced outlook. Readers are encouraged to review Part I carefully before proceeding.

Keywords: Groups, Rings, Fields, Polynomials, Roots of Polynomials, Matrices.

POLYNOMIALS

The concepts of polynomials and their operations, as taught in post-secondary education, require precise definition and rigorous justification. For example, it is commonly stated that since $x = 1$ is a root of the equation $x^3 - x^2 + x - 1 = 0$, the polynomial on the left-hand side can be divided by the binomial $x - 1$ without remainder. However, if x is interpreted strictly as an unknown satisfying the equation, this situation effectively represents a disguised form of division by zero, rendering the meaning of the operation ambiguous – despite the fact that the formal procedure of polynomial division proceeds without difficulty. Furthermore, in certain branches of mathematics, such as combinatorial analysis, one frequently encounters "polynomials" of the form

$$a_0 + a_1 x + a_2 x^2 + \cdots$$

with infinitely many nonzero coefficients. These objects, known as *formal power series*, are considered independently of the numerical values of x for which the partial sums

$$S_n(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

may converge as $n \to \infty$. Unlike analysis, where convergence is central, algebra focuses on the algebraic rules governing the formation of coefficients under polynomial operations. By introducing polynomials in a rigorous manner – including those with infinitely many terms – we can define addition and multiplication so that these operations align consistently with the rules of the standard "level of post-secondary education."

Let an arbitrary ring $K = (M, +, \cdot)$ be given. Consider the set M^∞, whose elements are all infinite sequences of elements from M. If $a, b \in M^\infty$, where

$$a = (a_0, a_1, a_2, \dots),\ b = (b_0, b_1, b_2, \dots),$$

then we define $a = b$ if and only if $a_i = b_i$ for every $i = 0, 1, 2, \dots$. Addition and multiplication in M^∞ are defined as follows:

$$a + b = (a_0 + b_0, a_1 + b_1, a_2 + b_2, \dots),$$
$$a \cdot b = (a_0 b_0,\ a_0 b_1 + a_1 b_0, a_0 b_2 + a_1 b_1 + a_2 b_0, \dots);$$

In general, the *j*-th term of the product $a \cdot b$ has the form

$$a_0 b_j + a_1 b_{j-1} + \cdots + a_j b_0 = \sum_{k=0}^{j} a_k b_{j-k} = \sum_{\substack{k+i=j \\ k,i \geq 0}} a_k b_i.$$

(In the last expression, it is tacitly assumed that the indices k and i take only those values satisfying the condition $k + i = j$ for a given j, where both k and i are nonnegative integers). Finally, note that it is convenient to refer to the *j*-th term of a sequence as what would ordinarily be the $(j + 1)$-th term in standard counting since here we begin indexing from zero rather than from one.

We now prove that the *system* $K^\infty = (M^\infty, +, \cdot)$ *is a ring*.

I. $(M^\infty, +)$ is an *abelian group*.

If $a, b \in M^\infty$, then the *j*-th term of $a + b$ is $(a + b)_j = a_j + b_j$ $(j = 0, 1, 2, \dots)$. Thus, all the axioms of an abelian group for sequences follow directly from the fact that $(M, +)$ is an abelian group since each coordinate operation is inherited termwise.

- The zero element of $(M^\infty, +)$ is the sequence $(0, 0, 0, \dots)$, where 0 is the additive identity of K.
- The additive inverse of $a = (a_0, a_1, a_2, \dots)$ is $-a = (-a_0, -a_1, -a_2, \dots)$.

Therefore, $(M^\infty, +)$ is an abelian group.

II. $(M^\infty, \cdot)$ is a *groupoid*.

If $a, b \in M^{\infty}$, then the *j*-th term of the product sequence ab is given by

$$(ab)_j = \sum_{k+i=j} a_k b_i,$$

which is an element of *M* because $K = (M, +, \cdot)$ is a ring. Hence $ab \in M^{\infty}$. Therefore, multiplication is closed in M^{∞}, and $(M^{\infty}, \cdot)$ is a groupoid.

III. *Distributive properties*.

Let $a = (a_0, a_1, a_2, \dots)$, $b = (b_0, b_1, b_2, \dots)$, $c = (c_0, c_1, c_2, \dots) \in M^{\infty}$. The *j*-th term of the sequence $(a + b)c$ is

$$\sum_{k+i=j} (a_k + b_k)c_i = \sum_{k+i=j} (a_k c_i + b_k c_i) = \sum_{k+i=j} a_k c_i + \sum_{k+i=j} b_k c_i .$$

This shows that the *j*-th term of $(a + b)c$ is equal to the sum of the corresponding *j*-th terms of ac and bc. Thus, $(a + b)c = ac + bc$. The proof of the other distributive property, $c(a + b) == ca + cb$, is analogous.

Theorem 15. *If a ring K has any of the properties*:

1. *commutativity*,
2. *associativity*,
3. *existence of a unit*,
4. *absence of zero divisors*,

then the ring of sequences K^{∞} also has the same properties.

Proof. 1. *Commutativity*.

Suppose *K* is commutative, i.e., $xy = yx$ for all $x, y \in M$. For any $j \geq 0$, the *j*-th term of the product sequence ab is

$$\sum_{k+i=j} a_k b_i = \sum_{i+k=j} b_i a_k,$$

which is exactly the *j*-th term of the sequence ba. Hence $ab = ba$ in K^{∞}.

2. *Associativity*.

Suppose *K* is associative. For any $j \geq 0$, the *j*-th term of the sequence $a(bc)$ is

$$\sum_{k+i=j} a_k \left(\sum_{p+q=i} b_p c_q \right) = \sum_{k+i=j} \sum_{p+q=i} a_k (b_p c_q) = \sum_{k+p+q=j} a_k (b_p c_q).$$

Similarly, the *j*-th term of the sequence $(ab)c$ is

$$\sum_{i+q=j} \left(\sum_{k+p=i} a_k b_p \right) c_q = \sum_{i+q=j} \sum_{k+p=i} (a_k b_p) c_q = \sum_{k+p+q=j} (a_k b_p) c_q.$$

Since multiplication in *K* is associative, we have $a_k (b_p c_q) = (a_k b_p) c_q$ for all k, p, q and therefore $a(bc) = (ab)c$ in K^{∞}.

3. *Existence of a unit*.

If 1 is the multiplicative identity in *K*, then for any sequence $(a_0, a_1, a_2, \dots) \in K^{\infty}$, we have

$$(1, 0, 0, \dots) \cdot (a_0, a_1, a_2, \dots) = (a_0, a_1, a_2, \dots) \cdot (1, 0, 0, \dots) = (a_0, a_1, a_2, \dots).$$

Hence the element $(1, 0, 0, \dots)$ serves as the multiplicative identity of the ring K^{∞}.

4. *Absence of zero divisors*.

Assume *K* has no zero divisors, and consider two arbitrary nonzero elements in K^{∞}:

$$a = (0, 0, \dots, 0, a_m, a_{m+1}, \dots) \text{ and } b = (0, 0, \dots, 0, b_n, b_{n+1}, \dots),$$

where $a_m \neq 0$ and $b_n \neq 0$ $(m, n \geq 0)$ are, respectively, the first nonzero terms of a and b. Then the $(m + n)$-th term of the product ab is

$$a_0 b_{m+n} + a_1 b_{m+n-1} + \cdots + a_{m-1} b_{n+1} + a_m b_n + a_{m+1} b_{n-1} + \cdots + a_{m+n} b_0.$$

Since $a_i = 0$ for $i < m$ and $b_k = 0$ for $k < n$, all terms vanish except $a_m b_n$. Thus,

$$(ab)_{m+n} = a_m b_n.$$

Because $a_m \neq 0$, $b_n \neq 0$, and *K* has no zero divisors, it follows that $a_m b_n \neq 0$. Therefore, $ab \neq (0, 0, 0, \dots)$, which shows that K^{∞} also has no zero divisors. •

Corollary 2. *If K is an integral ring, then so is K^{∞}*. •

However, if *K* is a field, then the ring K^{∞} (which is obviously nontrivial) is not a field. Indeed, the nonzero element $(0, 1, 0, 0, \dots)$ is not invertible: if there were some $b = (b_0, b_1, b_2, \dots)$ such that

$$(0, 1, 0, 0, \dots) \cdot b = (1, 0, 0, \dots),$$

it would follow that $0 \cdot b_0 = 1$, i.e., $0 = 1$, a contradiction.

Now let *K* be an associative ring with unity. Then so is K^{∞}. It is straightforward to show that the set of elements of the form $(a_0, 0, 0, \dots)$ constitutes a subring of K^{∞}, which is isomorphic to *K* via the natural isomorphism

$$f: K \longrightarrow K^{\infty}, f(a_0) = (a_0, 0, 0, \dots).$$

Accordingly, such an element can be denoted simply by a_0. In particular, $0 = (0, 0, 0, \dots)$ and $1 == (1, 0, 0, \dots)$ are, respectively, the zero and unity of the ring K^{∞}.

For the element $(0,1,0,0,\dots)$, we introduce the notation x. Then $(0,0,1,0,\dots) = x \cdot x = x^2$, the element $(0,0,0,1,0\dots) = x \cdot x^2 = x^2 \cdot x$ can be denoted by x^3, and so on. Assuming, by definition, that $x^0 = 1$, it is straightforward to prove that $x^m \cdot x^n = x^{m+n}$ for all integers $m, n \geq 0$. The element $(0,\dots,0,a_m,0,\dots)$ can be written in the form

$$(a_m,0,0,\dots)\cdot(0,\dots,0,1,0,\dots) = a_m x^m,$$

or equivalently, $(0,\dots,0,1,0,\dots)\cdot(a_m,0,0,\dots) = x^m a_m$. Hence, *every nonnegative integer power x^m commutes with all elements of the ring K^∞*, even though K^∞ itself need not be commutative.

Now consider an element $a = (a_0,a_1,a_2,\dots,a_m,0,0,\dots) \in M^\infty$, where $a_{m+1} = a_{m+2} = \cdots == 0$. It can be uniquely represented as

$$\begin{aligned} a = (a_0,0,0,\dots,0,0,\dots) + (0,a_1,0,\dots,0,0,\dots) + (0,0,a_2,\dots,0,0,\dots) + \cdots + (0,0,0,\dots,a_m,0,\dots) = \\ = (a_0,0,0,\dots) + (a_1,0,0,\dots)\cdot(0,1,0,0,\dots) + (a_2,0,0,\dots)\cdot(0,0,1,\dots) + \cdots + \\ +(a_m,0,0,\dots)\cdot(0,0,0,\dots,1,0,\dots) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m. \end{aligned}$$

That is, a is represented using the usual notation of a polynomial in terms of x. (One must prove the uniqueness of this representation.) Thus, every sequence in K^∞ with only finitely many nonzero terms corresponds to a polynomial, or more precisely – as is customary in higher algebra–to a *polynomial in a formal variable x over the ring K*. The word "formal" emphasizes that x is not just an element of K (or a "variable ranging over K"); rather, it is a well-defined, fixed element of the ring K^∞. The phrase "over the ring K" indicates that all coefficients lie in K. The subset of K^∞ consisting of such polynomials is denoted by $K[x]$.

Theorem 16. *The set $K[x]$ forms a subring of K^∞.*

Proof. (*Closure under addition*) Let

$$a = (a_0,a_1,\dots,a_m,0,0,\dots),\ b = (b_0,b_1,\dots,b_n,0,0,\dots) \in K[x].$$

Then their sum is $a + b = \left(a_0 + b_0, a_1 + b_1, \dots, a_{\max\{m,n\}} + b_{\max\{m,n\}}, 0, 0, \dots\right)$. Since all terms beyond index $\max\{m,n\}$ vanish, we conclude that $a + b \in K[x]$. Thus $(K[x],+)$ is a groupoid.

(*Closure under multiplication*) Consider the product ab. Since a has degree at most m and b has degree at most n, the degree of ab is at most $m+n$. Hence all terms beyond index $\max\{m,n\}$ vanish, and therefore $ab \in K[x]$. Thus $(K[x],\cdot)$is a groupoid.

(*Additive inverses*) For

$$a = (a_0,a_1,\dots,a_m,0,0,\dots) \in K[x],$$

the additive inverse is $-a = (-a_0,-a_1,\dots,-a_m,0,0,\dots) \in K[x]$.

Therefore, $K[x]$ is closed under addition, multiplication, and contains additive inverses. Since it is already a subset of the ring K^∞, we conclude that $K[x]$ is a subring of K^∞. •

Remark 17. As we already know, these three properties are sufficient for $K[x]$ to be a subring of K^∞. On this basis, $K[x]$ is called the *ring of polynomials in x over the ring K*. Elements of $K[x]$ will be written in polynomial form rather than as sequences.

Indeed, the coefficients of the powers $x^0 = 1, x^1 = x, x^2, \dots, x^m$ in the polynomial

$$A(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m$$

correspond to the first $m+1$ terms of the sequence

$$(a_0,a_1,\dots,a_m,0,0,\dots),$$

which represents the same element (but in different notation) of the subring $K[x]$. Since two sequences in K^∞ are equal if and only if their corresponding terms are equal, it follows that *two polynomials are equal – that is, they represent the same element of $K[x]$ – if and only if their coefficients at each power of the formal variable coincide*.

Polynomials from $K[x]$ will be denoted by $A(x), B(x), C(x)$, ..., or simply $A, B, C, \dots$ in capital letters, so as not to confuse them with elements of the ring K.

If $A = A(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m \in K[x]$, $a_m \neq 0, m \geq 0$, then the number m is called the *degree* of the polynomial A, denoted $\deg A$. In particular:

- If $\deg A = 0$, then $A = a_0$ contains no variable x, and thus corresponds to a nonzero element of K.
- For $a_0 = 0$, the zero polynomial $A = 0$ has no defined degree; nevertheless, it is convenient to assign it a negative degree (without specifying which).

Theorem 17. *Let $A, B \in K[x]$. Then*:

1. $\deg(A+B) \leq \max\{\deg A, \deg B\}$;
2. $\deg(A \cdot B) \leq \deg A + \deg B$.

Moreover, if the leading coefficients $a_m \neq 0$ and $b_n \neq 0$ of the polynomials A and B are not zero divisors in K, then equality holds in both cases. •

Remark 18. The first statement is straightforward. Note, however, that the degree of the sum of two polynomials may be strictly less than the degree of each individual polynomial, even when $\deg A = \deg B$. This occurs precisely when the leading coefficients (i.e., the coefficients corresponding to the highest power of x) of A and B are additive inverses in the ring K. To prove the second statement, let

$$A = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m, \text{ where } a_m \neq 0,\ m = \deg A,$$
$$B = b_0 + b_1x + b_2x^2 + \cdots + b_nx^n, \text{ where } b_n \neq 0,\ n = \deg B.$$

then the highest possible power of x in the product $A \cdot B$ is $m+n$ and the leading term is $a_m b_n x^{m+n}$. This term is not zero whenever a_m and b_n are not zero divisors in K.

The inequalities (1) and (2) of Theorem 17 also remain valid when $A = 0$ or $B = 0$, due to the convention that $\deg(0) < 0$.

DIVISION WITH REMAINDER IN $K[x]$. We continue to assume that the ring K is associative with unity 1; hence, the same is true for $K[x]$. From now on, it will be more convenient to write polynomials in $K[x]$ in *descending* order of powers of x, rather than ascending order.

Theorem 18. *Let*

$$A = a_m x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + a_0,\ B = b_n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0$$

be two polynomials in $K[x]$, where $B \neq 0$ and the leading coefficient b_n has an inverse element b_n^{-1} in the ring K. Then:

1. *There exist unique polynomials $Q_{left}, R_{left} \in K[x]$ such that*

$$A = Q_{left} \cdot B + R_{left},\ \deg R_{left} < \deg B.$$

2. *There exist unique polynomials $Q_{right}, R_{right} \in K[x]$ such that*

$$A = B \cdot Q_{right} + R_{right},\ \deg R_{right} < \deg B.$$

Proof. The polynomials Q_{left} and R_{left} are called the *left quotient* and the *left remainder* when dividing A (*the dividend*) by B (*the divisor*). Similarly, Q_{right} and R_{right} are called the *right quotient* and the *right remainder*. We establish their existence and uniqueness separately.

(*Left division*) The existence of such polynomials is established by following the usual scheme of polynomial long division, but with careful attention to the difference between left and right division in the case when the coefficient ring K is noncommutative. We begin with the left division.

$$\begin{array}{l} a_m x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + a_0 \quad \Big| \quad b_n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0 \\ \hline a_m x^m + a_m b_n^{-1} b_{n-1} x^{m-1} + \cdots \qquad\qquad a_m b_n^{-1} x^{m-n} + (a_{m-1} - a_m b_n^{-1} b_{n-1}) b_n^{-1} x^{m-n-1} + \cdots \\ \hline (a_{m-1} - a_m b_n^{-1} b_{n-1}) x^{m-1} + \cdots \end{array}$$

If $\deg A < \deg B$, then we immediately set $Q_{left} = 0$, $R_{left} = A$. If $\deg A \geq \deg B$, then we begin by choosing the leading term of the left quotient Q_{left} as $a_m b_n^{-1} x^{m-n}$, so that when multiplied on the left by B, the leading term matches that of A:

$a_m b_n^{-1} x^{m-n} \cdot (b_n x^n + \cdots) = a_m b_n^{-1} x^{m-n} \cdot b_n x^n + \cdots = a_m b_n^{-1} (x^{m-n} b_n) x^n + \cdots$
$= a_m b_n^{-1} (b_n x^{m-n}) x^n + \cdots = a_m (b_n^{-1} b_n)(x^{m-n}) x^n + \cdots = a_m \cdot 1 \cdot x^{m-n+n} = a_m x^m + \cdots$

Subtracting this from A, we obtain

$$A - (a_m x^m + a_m b_n^{-1} b_{n-1} x^{m-1} + \cdots) = (a_{m-1} - a_m b_n^{-1} b_{n-1}) x^{m-1} + \cdots,$$

which is a polynomial of degree strictly less than m (equal to $m-1$ if $a_{m-1} - a_m b_n^{-1} b_{n-1} \neq 0$, otherwise even smaller).

If the degree of this remainder is already less than $\deg B$, then we take it as R_{left} and set

$$Q_{left} = a_m b_n^{-1} x^{m-n}.$$

If instead the degree of the remainder is still at least $\deg B$, we proceed to determine the next term of the quotient. This second term (of degree less than $m-n$) is chosen so that, after multiplying the sum of the two first quotient terms by B, the resulting product agrees with A in both the leading term $a_m x^m$ and the next term $a_{m-1} x^{m-1}$.

This process is then continued inductively. At each step, the degree of the current remainder strictly decreases, ensuring that after finitely many steps we obtain a remainder of degree less than $\deg B$. We take this polynomial as R_{left}, while Q_{left} is defined as the sum of all the monomials chosen in the successive steps. Thus,

$$A = Q_{left} \cdot B + R_{left},\ \deg R_{left} < \deg B.$$

This description also covers the case of exact division, where $R_{left} = 0$. In this case we note that $\deg(0) = -\infty < \deg B$, and since $b_n \neq 0$ (due to its invertibility), we have $\deg B \geq 0$.

(*Right division*) The existence of Q_{right} and R_{right} is proved in a similar way (they do not necessarily coincide with Q_{left} and R_{left}), except that the quotient terms are chosen to act on the right of B. Equivalently, we select the quotient coefficients as indicated and multiply them on the left of B.

$$\begin{array}{l} a_m x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + a_0 \quad \Big| \quad b_n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0 \\ \hline a_m x^m + b_{n-1} b_n^{-1} a_m x^{m-1} + \cdots \qquad\qquad b_n^{-1} a_m x^{m-n} + b_n^{-1} (a_{m-1} - b_{n-1} b_n^{-1} a_m) x^{m-n-1} + \cdots \\ \hline (a_{m-1} - b_{n-1} b_n^{-1} a_m) x^{m-1} + \cdots \end{array}$$

The uniqueness of the quotient and the remainder in both cases is proved in a completely similar way and we now consider the left division (omitting the index "left" for simplicity). Let two representations be obtained as follows:

$$A' = Q \cdot B + R = \tilde{Q} \cdot B + \tilde{R},\ \deg R < \deg B,\ \deg \tilde{R} < \deg B;$$

then $(Q-\tilde{Q})\cdot B=\tilde{R}-R$. For $Q-\tilde{Q}\neq 0$, we would have $\deg(Q-\tilde{Q})\geq 0$ and since the polynomial B has leading coefficient b_n, which is not a zero divisor in the ring K because it is invertible, $\deg\left((Q-\tilde{Q})\cdot B\right)=\deg(Q-\tilde{Q})+\deg B\geq \deg B$ while $\deg(\tilde{R}-R)\leq$

$\leq \max\{\deg\tilde{R},\deg(-R)\}<\deg B$, which is not possible. Therefore, $\tilde{Q}=Q$. It follows that $\tilde{R}-R=0\cdot B=0$, i.e., $\tilde{R}=R$. This completes the proof. •

If the ring K (and therefore $K[x]$) is commutative, then $Q_{left}=Q_{right}$ and $R_{left}=R_{right}$. In this case, the quotient and the remainder are simply denoted by Q and R. The following is a restatement of Theorem 18.

Theorem 18. (*Restated*) *For any polynomials $A,B\in K[x]$, where the leading coefficient b_n of B is invertible, there exist unique polynomials $Q,R\in K[x]$ such that*

$$A=QB+R=BQ+R,\ \deg R<\deg B.$$

If $R=0$, then B is called a *divisor* of A. In this case, we also say that B *divides* A or that A *is divisible by* B.

DIVISIBILITY OF POLYNOMIALS OVER A FIELD

Let's study in detail the divisibility properties of polynomials under the assumption that the base ring K is a field P. Since every field is an integral ring, the ring $P[x]$ is also integral. For any $A,B\in P[x]$, the following properties hold:

$$\deg(A\pm B)\leq\max\{\deg A,\deg B\},$$
$$\deg(A\cdot B)=\deg A+\deg B.$$

Lemma 1. *Let $A,B,C,A_1,A_2,\dots,A_k,C_1,C_2,\dots,C_k\in P[x]$. Then*:

1. *If A is divisible by B, and B is divisible by C, then A is divisible by C.*
2. *If $A_1,A_2,\dots,A_k$ are each divisible by B, and $C_1,C_2,\dots,C_k$ are arbitrary, then*

$$A_1C_1+A_2C_2+\cdots+A_kC_k$$

is divisible by B. In particular, $A_1\pm A_2$ and A_1C_1 are divisible by B.

Proof.

1. Divisibility of A by B means that there exists $Q_1\in P[x]$ such that $A=BQ_1$. Divisibility of B by C means that there exists $Q_2\in P[x]$ such that $B=CQ_2$. Then $A=BQ_1=CQ_2\cdot Q_1=C\cdot Q_2Q_1$. Setting $Q_3=Q_2Q_1\in P[x]$, we get $A=CQ_3$. Hence, A is divisible by C.

2. If $A_1,A_2,\dots,A_k$ are divisible by B, then there exist $Q_1,Q_2,\dots,Q_k\in P[x]$ such that $A_1=BQ_1$, $A_2=BQ_2$,…, $A_k=BQ_k$. Then

$$A_1C_1+A_2C_2+\cdots+A_kC_k=BQ_1C_1+\ BQ_2C_2+\cdots+BQ_kC_k=BQ,$$

where $Q=Q_1C_1+Q_2C_2+\cdots+Q_kC_k\in P[x]$. Therefore, $A_1C_1+\ A_2C_2+\cdots+A_kC_k$ is divisible by B. •

Lemma 2. *Let $A,B\in P[x]$. The divisibility* (*or indivisibility*) *of A by B is preserved when both polynomials are multiplied by any nonzero elements of the field P.*

Proof. Suppose A is divisible by B. Then there exists $Q\in P[x]$ such that $A=BQ$. Let $c,d\in P\setminus\{0\}$. Then $cA=c(BQ)=dB\cdot cd^{-1}Q$, and since $d^{-1}\in P$, it follows that $cd^{-1}Q\in P[x]$. Thus, cA is divisible by dB. Conversely, suppose cA is divisible by dB. Then, there exists $\tilde{Q}\in P[x]$ such that $cA=\ dB\cdot\tilde{Q}$. Multiplying both sides by c^{-1} (which exists since $c\neq 0$) gives $A=B\cdot c^{-1}d\tilde{Q}$, and $c^{-1}d\tilde{Q}\in P[x]$. Hence, A is divisible by B. •

Corollary 3.

1. *Every $A\in P[x]$ is divisible by every $B\in P[x]$ with degree* $\deg B=0$; *that is, by every nonzero element of the field P. In other words, dividing a polynomial A by a nonzero constant rescales the coefficients of A and the quotient remains a polynomial in $P[x]$.*
2. *If* $\deg A=0$ *and A is divisible by B, then* $\deg B=0$. •

Remark 19. Unfortunately, pages 67 *and* 68 *of the original Russian Version of this algebra course are missing. However, the subsequent exposition appears to be largely unaffected by this omission. We apologize for any inconvenience this may cause.*

Remark 20. It is our view that the reconstructed material corresponding to the missing pages 67 and 68 can be presented in the following general form:

Theorem. (*Division Algorithm for Polynomials*) *Let $A,B\in P[x]$ with $B\neq 0$. Then there exist unique $Q,R\in P[x]$ such that*

$$A=BQ+R,\ \deg R<\deg B$$

Definition. A polynomial $D\in P[x]$ is called the *greatest common divisor* of $A,B\in P[x]$, denoted by $\gcd(A,B)$, if

1. both A and B are divisible by D;
2. D is divisible by any other common divisor of A and B.

By convention, the gcd is chosen *monic* (leading coefficient 1).

Lemma 3. *If D and D_1 are both* $\gcd(A,B)$, *then $D=cD_1$ for some $c\in P\setminus\{0\}$. In particular, the monic* gcd *is unique*.

Comment. Existence of gcds follows from the Euclidean algorithm for polynomials: repeated multiplication of the *Division Algorithm* yields the last nonzero remainder, which is the gcd of the two polynomials. •

Theorem-Lemma 1. *For any two polynomials $A,B\in P[x]$, with at least one of them nonzero, there exist $U,V\in P[x]$ such that $UA+VB=\gcd(A,B)$.* •

Remark 21. This result follows directly from the Euclidean algorithm: we express R_k in terms of R_{k-1} and R_{k-2}, then express R_{k-1} in terms of R_{k-2} and R_{k-3}, and so forth, until R_k is represented solely in terms of A and B. Subsequently, we multiply the polynomials that are factors of A and B by r^{-1}, where r is the leading coefficient of the last nonzero remainder R_k. This normalization converts R_k into a *monic polynomial* – a polynomial whose leading coefficient is 1. To avoid unnecessary complexity, we will demonstrate this process with a numerical example.

Example 24. Let $A = 64$, $B = 28$. Then

$$\begin{array}{rr|l} 64 & 28 & \\ \underline{56} & 2 & \\ 28 & 8 & \\ \underline{24} & 3 & \\ 8 & 4 & \\ \underline{8} & 2 & \\ 0 & & \end{array}$$

$64 = 28(2) + 8,\ \ 8 = 64 - 2(28),$
$28 = 8(3) + 4,\ \ \ \ 4 = 28 - 3(8)$
$8 = 4(2);$
$4 = 28 - 3(8) = 28 - 3[64 - 2(28)]$
$= 28 - 3(64) + 6(28) = -3(64) + 7(28);$
$\gcd(64, 28) = 4;$
here $U = -3, V = 7.$

It is not difficult to show in general that such a representation of $\gcd(A, B)$ is not unique (the proof is left to the reader).

Polynomials $A, B \in P[x]$ are called *coprime* if $\gcd(A, B) = 1$.

Corollary 4. (*from Theorem-Lemma 1*) *A and B are coprime if and only if there exist U and B such that $UA + VB = 1$.*

Proof. Statement "⇒" is a special case of Theorem-Lemma 1. Let's prove "⇐". Let there be such U and B. Then, by Lemma 1, polynomial 1 is divisible by any $cd(A, B)$ and by Corollary 3, the degree of it is 0, in particular, $\deg \gcd(A, B) = 0$. Hence, $\gcd(A, B) = 1$ due to the fact that this polynomial is monic. •

Theorem 19. *If* $\gcd(A, C) = 1$, *then* $\gcd(A, BC) = \gcd(A, B)$. ($A, B, C \in P[x]$; $A \neq 0$, $C \neq 0$.)

Proof. By Corollary 4, there exist $U_1, V_1 \in P[x]$ such that

$$U_1 A + V_1 C = 1,$$

and, by Theorem-Lemma 1 itself, there exist $U_2, V_2 \in P[x]$ such that

$$U_2 A + V_2 B = \text{gdc}(A, B);$$

multiplying these two equalities, we get

$$U \cdot A + V \cdot BC = \gcd(A, B),$$

where $U = U_1 U_2 A + U_1 V_2 B + V_1 U_2 C \in P[x], V = V_1 V_2 \in P[x]$. From here, by Lemma 1, it follows that $\gcd(A, B)$ is divisible by any $cd(A, B)$, and therefore also by $\gcd(A, BC)$. Conversely, A and BC are divisible by any $cd(A, B)$, in particular, by $\gcd(A, B)$, therefore $\gcd(A, BC)$ is also divisible by $\gcd(A, B)$. Since both greatest common divisors are monic polynomials, it follows from Lemma 3 (which is missing in the original text but has been reconstructed in Remark 20) that $\gcd(A, BC) = \gcd(A, B)$. •

Corollary 5. *If* $\gcd(A, B) = \gcd(A, C) = \gcd(A, D) = \cdots = 1$, *then* $\gcd(A, BCD \ldots) = 1$; *in other words, if a polynomial A is coprime with each of the polynomials $B, C, D, \ldots$, then it is coprime with their product*. •

Theorem 20. *If A is divisible by B and C, and* $\gcd(B, C) = 1$, *then A is divisible by BC*. ($A, B, C \in P[x]$; $B \neq 0$, $C \neq 0$.)

Proof. Under the given conditions, there exists $Q \in P[x]$ such that $A = BQ$. Moreover, by Corollary 4, there exist $U, V \in P[x]$ such that $UB + VC = 1$. Multiplying the last equality by Q, we obtain

$$UBQ + VCQ = Q.$$

Since polynomial $BQ = A$ is divisible by C under the given conditions, Lemma 1 implies that Q is divisible by C; that is, there exists $D \in P[x]$ such that $Q = CD$. Hence,

$$A = BQ = B \cdot CD = BC \cdot D,$$

i.e., A es divisible by BC. •

REDUCIBILITY OF POLYNOMIALS

A polynomial $A \in P[x]$, $A \neq 0$ (i.e., $\deg A \geq 0$) is called *reducible over a field P* (or *in the ring $P[x]$*) if there exist polynomials $B, C \in P[x]$ such that $A = BC$, and

$0 < \deg B < \deg A$, $0 < \deg C < \deg A$. (3)

In other words, A is reducible if it can be expressed as the product of two polynomials over the same field P with positive degrees. Since $\deg A = \deg B + \deg C$, each of the inequalities in (3) implies the other. It follows directly from this definition that *polynomials of degree* 0 *and degree* 1 *are irreducible*.

If a polynomial $A \in P[x]$ (that is, over the field P) is irreducible over P and $\hat{P}$ is an extension of P (i.e., $P \subset \hat{P}$ and P is a subfield of $\hat{P}$), then $A \in \hat{P}[x]$ (a polynomial over a given field is also a polynomial over any of its extensions). However, over the extended field $\hat{P}$ the same polynomial may become reducible. For example, let $P = \mathbb{Q}$ and $\hat{P} = \mathbb{Q}(\sqrt{2})$. The polynomial $A = x^2 - 2$ belonging to $\mathbb{Q}[x]$ (and therefore to $\mathbb{Q}(\sqrt{2})[x]$) is irreducible over P, but reducible over $\hat{P}$. Indeed, if we try to factor $x^2 - 2$ as $(x + a)(x + b)$ with $a, b \in \mathbb{Q}$, and writing the equality in the form $x^2 + 0 \cdot x + (-2) = x^2 + (a + b) \cdot x + ab$, we will have $a + b = 0$ and $ab = -2$, which makes $a^2 + ab = 0$ and $ab = -2$. This implies $a^2 = 2$, which has no solution in $\mathbb{Q}$. However, over $\mathbb{Q}(\sqrt{2})$, the

factorization $x^2-2=\left(x+\sqrt{2}\right)\left(x-\sqrt{2}\right)$ is valid. Another example is x^2+1. Over $\mathbb{Q}$, $\mathbb{Q}\left(\sqrt{2}\right)$ and $\mathbb{R}$ this polynomial is irreducible, but over $\mathbb{C}$ it factors as $x^2+1=(x+i)(x-i)$. (You can verify the irreducibility of this polynomial over $\mathbb{R}$ yourself).

Remark 22. If polynomials $A,B\in P[x]$ have a common divisor *cd* over some extension of the field P, it is not necessary that $cd\in P[x]$. However, the greatest common divisor $\gcd(A,B)$ always belongs to $P[x]$. This is because, when computing the gcd using the Euclidean algorithm, only operations on the coefficients that remain within the field P (addition, subtraction, multiplication, and division) are performed. For example, polynomials

$$A=x^5-x^3+x^2-2x-2,\ B=2x^3+x^2-4x-2\in\mathbb{Q}[x]$$

have a common divisor $cd(A,B)=x-\sqrt{2}\notin\mathbb{Q}[x]$, but their greatest common divisor satisfies $\gcd(A,B)=x^2-2\in\mathbb{Q}[x]$. It follows directly from this remark

Lemma 4. *If monic polynomials $A,B\in P[x]$ are irreducible over the field P, then either $A=B$ or* $\gcd(A,B)=1$. •

Remark 23. Let's also recall that, due to the absence of zero divisors in the ring $P[x]$, if $AC=BC$ with $C\neq 0$, then $A=B$; that is, both sides of the equality can be divided by a common factor of nonzero degree. This is precisely Theorem 19 on rings applied to $P[x]$.

Theorem 21. *Any polynomial $S\in P[x]$ can be expressed in the form $S=aA_1A_2\dots A_k$, where $a\in P$ is the leading coefficient of S, and each $A_i\in P[x]$* for $i=1,2,\dots,k$ *is monic and irreducible over the field P. Note that the factors $A_1,A_2,\dots,A_k$ may include repetitions. For $S\neq 0$, this factorization is unique up to the order of the factors.*

Proof. For $S=0$, we may, for instance, set $a=0$ and choose any collection of monic irreducible polynomials $A_1,A_2,\dots$ in any quantity. In this case, uniqueness of the factorization clearly does not hold.

Let $S\neq 0$. If S is already monic, we set $a=1$; otherwise, we write S in the form $S=a\cdot a^{-1}S$, where a is the leading coefficient of S, and $a^{-1}S$ is monic (the element $a\in P$ is invertible since $a\neq 0$). If this polynomial is irreducible over P, then, setting $a^{-1}S=A_1$, gives the desired factorization $S=aA_1$ (with $k=1$). In the case where the polynomial $a^{-1}S$ is reducible, we factor it into a product of two polynomials of lower degree over the same field P. If at least one of these factors is itself reducible, we factor it further and continue this process iteratively. These factorizations can always be performed so that all factors are monic. Since each factorization produces polynomials of strictly smaller degree, and polynomials of degree 1 are irreducible, the process must terminate after finitely many steps, yielding the expression

$S=aA_1A_2\dots A_k$, (4)

where $a\in P$ and polynomials $A_1,A_2,\dots,A_k$ are monic and irreducible over P, and

$$\deg A_1+\deg A_2+\dots+\deg A_k=\deg S,$$

$k\geq 0$ ($k=0$ if $\deg S=0$, i.e., $S=a$). We will prove the uniqueness of expression (4).

Assume that for some polynomial $S\in P[x]$, there exist two different factorizations satisfying the theorem conditions:

$$S=aA_1A_2\dots A_k=bB_1B_2\dots B_l,$$

where $a\neq 0$ and $b\neq 0$ since $S\neq 0$. In fact, $b=a$, because both a and b are the leading coefficients of S, and all remaining factors in this factorization are monic (assuming $k>0$ and $l>0$). Multiplying both sides of the equality by a^{-1}, we obtain

$A_1A_2\dots A_k=B_1B_2\dots B_l$. (5)

The Polynomial A_1 cannot be coprime with each of the polynomials $B_1,B_2,\dots,B_l$; if it were, then by Corollary 5, it would be coprime with their product. However, the product on the right-hand side of (5) is divisible by A_1 (since $\deg A_1>0$) because both sides of equality (5) represent the same polynomial. Hence, there exists some B_j ($1\leq j\leq l$) such that $\gcd(A_1,B_j)\neq 1$. By Lemma 4, since A_1 and B_j are monic and irreducible, it follows that $A_1=B_j$. Dividing both sides of (5) by A_1 on the left and by B_j on the right (see Remark 23), we obtain

$$A_2\dots A_k=B_1B_2\dots B_{j-1}B_{j+1}\dots B_l.$$

Applying the same reasoning to the new equality, we identify a factor on the right-hand side equal to A_2, cancel it from both sides, and continue in this manner until one side becomes 1, meaning all factors have been exhausted. Since both sides are monic polynomials of the same degree, the other side must be also 1. Consequently, $l=k$, and the polynomials $A_1,A_2,\dots,A_k$ coincide with $B_1,B_2,\dots,B_k$, though not necessarily in the same order. The statement now follows. •

By grouping identical irreducible factors in (4) (renumbering them accordingly and letting k denote the number of distinct factor groups rather than the total number of factors), we can write the polynomial $S\in P[x]$ in the form

$$S=aA_1^{r_1}\cdot A_2^{r_2}\cdot\dots\cdot A_k^{r_k},$$

where S is an arbitrary polynomial over the field P, $a\in P$, and $A_1,A_2,\dots,A_k$ are monic, irreducible polynomials over P, pairwise coprime (and thus distinct). In this case,

$$r_1\deg A_1+r_2\deg A_2+\dots+r_k\deg A_k=\deg S.$$

The natural number r_j is called the *multiplicity* of the factor A_j in the product S ($j=1,2,\dots,k$ with $k\geq 0$). The representation is unique up to a permutation of the factors $A_j^{r_j}$ among themselves.

ASSIGNMENT FOR INDEPENDENT WORK.

(*based on A.G. Kurosh, Course of Higher Algebra*).

1. Give a purely algebraic definition of the derivative of a polynomial. Derive the rules for differentiating the sum and the product of polynomials, and explain why results from mathematical analysis cannot be directly applied here.
2. Let P be a field of characteristic $p = 0$, and let
$$S = aA_1^{r_1}A_2^{r_2} \dots A_k^{r_k}$$
be the factorization of $S \in P[x]$ into irreducible factors. Prove that
$$\frac{S}{\gcd(S,S')} = aA_1A_2 \dots A_k,$$
that is, dividing S by the greatest common divisor of S and its derivative S' yields a polynomial containing the same irreducible factors as S, but each to the first power.
3. Investigate the role of the condition $p = 0$.

VALUES AND ROOTS OF POLYNOMIALS

We already know that in the expression
$$A(x) = a_nx^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0$$
for the polynomial $A = A(x) \in P[x]$, the formal variable x is defined as the element $(0, 1, 0, \dots)$ of the ring of polynomials over the field P, and not as "a blank space to be filled with an arbitrary element of P". Nevertheless, the operation of *substituting x with an element $c \in P$ in the polynomial $A(x)$* can be rigorously defined by stipulating that the result of such a substitution is the element
$$A(c) = a_nc^n + a_{n-1}c^{n-1} + \dots + a_1c + a_0 \in P$$
obtained from c by performing the same operations (in the same order) by which $A(x)$ is obtained from x. This resulting element $A(c)$ is called *the value of the polynomial $A(x)$ at $x = c$*. If A and B are the same polynomial (written as $A = B$ or $A(x) = B(x)$), then clearly $A(c) = B(c)$ for any $c \in P$. However, the converse does not hold for every field P. For example, over the field $\mathbb{Z}_2$ (of characteristic $p = 2$) consisting of two elements $\{0, 1\}$–where $1 + 1 = 0$ – the polynomials $A(x) = x + 1, B(x) = x^2 + 1$ are distinct yet satisfy $A(0) = B(0) = 1$ and $A(1) = B(1) = 0$. We will not generalize this example here, but note the underlying reason for the phenomenon: the field $\mathbb{Z}_2$ contains so few elements that the variety of distinct polynomials over it is much greater than the number of possible function values.

An element c is called a *root* of a polynomial $A(x)$ (or of the equation $A(x) = 0$) if $A(c) = 0$. The element c need not to belong to the field P over which the polynomial A is defined, it may lie in some extension $\hat{P}$ of P. The definition of $A(c)$ remains valid in this case since $A(x) \in P[x]$ implies $A(x) \in \hat{P}[x]$. For example, the polynomial $x^2 - 2$ over the field of rational numbers has no roots in $\mathbb{Q}$, but has roots in its extension $\mathbb{Q}(\sqrt{2})$ as well as in larger (nonminimal) extensions containing it, such as $\mathbb{Q}(\sqrt{2}, \sqrt{3})$, $\mathbb{R}$ and others.

Theorem 22. (*Little Bézout's Theorem, also known as the Polynomial Remainder Theorem*) *The remainder upon dividing a polynomial $A(x) \in P[x]$ by a monic first-degree binomial $x - c$ (which reduces to a monomial when $c = 0$), where $c \in P$, is equal to the value $A(c)$. In particular, $A(x)$ is divisible by $x - c$ if and only if c is a root of $A(x)$.*

Proof. When dividing $A(x)$ by $x - c$ with remainder, we have
$$A(x) = (x - c)Q(x) + R(x),$$
where $\deg R(x) < \deg(x - c)$. Moreover, both the remainder $R(x)$ and the quotient $Q(x)$ are uniquely determined by $A(x)$ and c. Since $\deg(x - c) = 1$, it follows that $\deg R(x) \le 0$; that is, $R(x)$ is constant and therefore an element of the field P. Thus, the equality
$$A(x) = (x - c)Q(x) + R \quad (6)$$
becomes an equality of elements of P when x is replaced by any element of P. In particular, substituting $x = c$ yields $A(c) = (c - c)Q(c) + R = R$. •

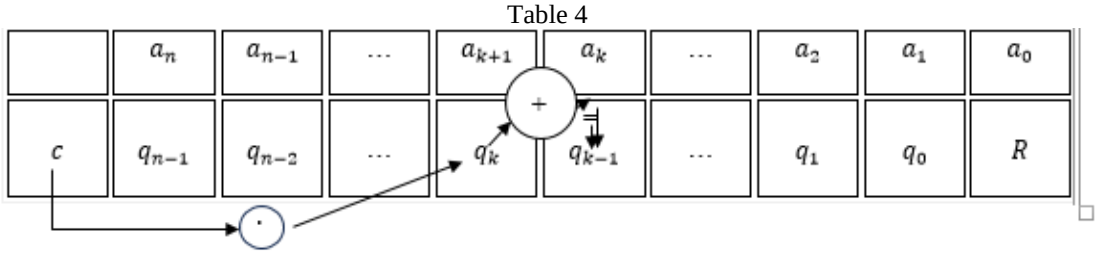
Thanks to Theorem 22, we can determine whether a polynomial $A(x)$ is divisible by a binomial $x - c$ without explicitly performing the division. Moreover, there exists an efficient procedure to compute both the remainder R and the coefficients of the quotient
$$Q(x) = q_{n-1}x^{n-1} + q_{n-2}x^{n-2} + \dots + q_1x + q_0,$$
requiring far fewer multiplications than directly substituting c into $A(x)$. This method is traditionally attributed to William Horner (18th-19th centuries), although historical evidence indicates that it was already known in China well before the Common Era.

Since the equality of two polynomials implies that the coefficients of corresponding powers of the formal variable x are identical, expanding the brackets on the right-hand side of the fully written equality (6)
$$a_nx^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0 = (x - c)(q_{n-1}x^{n-1} + q_{n-2}x^{n-2} + \dots + q_1x + q_0) + R$$
and by equating the corresponding coefficients, we obtain

$$\begin{aligned} a_n &= q_{n-1}, \\ a_{n-1} &= q_{n-2} - cq_{n-1}, \\ a_{n-2} &= q_{n-3} - cq_{n-2}, \\ &\dots\ \dots\ \dots \\ a_2 &= q_1 - cq_2, \\ a_1 &= q_0 - cq_1, \\ a_0 &= R - cq_0, \end{aligned} \qquad \text{where} \qquad \begin{aligned} q_{n-1} &= a_n, \\ q_{n-2} &= cq_{n-1} + a_{n-1}, \\ q_{n-3} &= cq_{n-2} + a_{n-2}, \\ &\dots\ \dots\ \dots \\ q_1 &= cq_2 + a_2, \\ q_0 &= cq_1 + a_1, \\ R &= cq_0 + a_0. \end{aligned}$$

The sequential computation of q_{n-1}, q_{n-2},…, q_1, q_0, R can be conveniently represented using *Horner's scheme* (Table 4).

Table 4



Example 25. Divide $A(x) = 2x^7 + 15x^6 + 240x^4 - 27x^3 + 10x^2 + 94x + 36$ by $x + 9$ and determine whether $A(-9) = 0$.

In this case, $c = -9$. After entering the coefficients of the polynomial $A(x)$ in the top row of Table 5, fill in the cells of the bottom row sequentially from left to right.

Table 5

	2	15	0	240	-27	10	94	36
-9	2	-3	27	-3	0	10	4	0

Remark 24. The coefficients of the last row of Table 5 (computed using Horner's scheme) can equivalently be obtained by applying *synthetic division* to the coefficients of $A(x)$ in the usual way:

2	15	0	240	-27	10	94	36	-9
	-18	27	-243	27	0	-90	-36	
2	-3	27	-3	0	10	4	0	

Thus, $Q(x) = 2x^6 - 3x^5 + 27x^4 - 3x^3 + 10x + 4$, $R = 0$, i.e., $A(x)$ is divisible by $x + 9$, and therefore $A(-9) = 0$. If we attempt to verify directly whether -9 is a root of $A(x)$, the calculation would require $8 + 7 + 5 + 4 + 3 + 1 = 28$ multiplications. In contrast, using Horner's scheme the multiplication operation is performed only 7 times (including one multiplication by zero).

From Theorem 22 it follows that *the zero element* ($c = 0$) *of the field P is a root of* $A(x) \in P[x]$ *if and only if* $a_0 = 0$. In this case, Horner's scheme becomes trivial: to compute the quotient $A(x)/x$, it suffices to reduce each power of x by 1 in $A(x)$.

TASK (optional): Describe and justify two variants – "right" and "left" – of Horner's schemes for polynomials over a noncommutative (but associative) ring with unity.

ABSOLUTE REDUCIBILITY OF POLYNOMIALS OVER THE FIELD $\mathbb{C}$. By the Fundamental Theorem of Algebra, any polynomial $A(x)$ of positive degree with complex coefficients has at least one complex root $c_1 \in \mathbb{C}$ *(The theorem does not apply to a polynomial of degree* 0 *or negative since in that case* $A(x) = 0$ *is not an equation)*. By Theorem 22, $A(x)$ is divisible by the binomial $x - c_1$; that is, there exists a polynomial $Q_1(x) \in \mathbb{C}[x]$ such that $A(x) = (x - c_1)Q_1(x)$. If $\deg Q_1 > 0$, then $Q_1(x)$ also has a root $c_2 \in \mathbb{C}$ (possibly equal to c_1), so that $Q_1(x) = (x - c_2)Q_2(x)$ and hence $A(x) = (x - c_1)(x - c_2)Q_2(x)$. Continuing this process, we obtain a sequence of polynomials $A, Q_1, Q_2, \ldots$ each having degree one less than its predecessor. After n steps, we arrive at a constant $Q_n = a_n \in \mathbb{C}$**,** which is the leading coefficient of the original polynomial $A(x)$ *(why?)*. Thus, we obtain the factorization – unique up to permutation of the linear factors:

$A(x) = a_n(x - c_1)(x - c_2) \ldots (x - c_n)$. (7)

Since polynomials of degree zero or one are irreducible over any field, a decomposition of the form (7) represents an *extremely fine level of factorization*. For this reason, we say that polynomials over $\mathbb{C}$ are *absolutely reducible* in this field (or in the ring $\mathbb{C}[x]$).

If $A(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, then by expanding the right-hand side of equality (7) and equating the coefficients of corresponding powers of x, we obtain Vieta's formulas

$$a_n \cdot \sum_{j_1=1}^{n} c_{j_1} = -a_{n-1},$$

$$a_n \cdot \sum_{1 \le j_1 < j_2 \le n}^{n} c_{j_1} c_{j_2} = a_{n-2},$$

$$a_n \cdot \sum_{1 \le j_1 < j_2 < j_3 \le n}^{n} c_{j_1} c_{j_2} c_{j_3} = -a_{n-3},$$

$$\cdots\cdots\cdots\cdots$$

$$a_n \cdot \sum_{1 \le j_1 < j_2 < \ldots < j_n \le n}^{n} c_{j_1} c_{j_2} \ldots c_{j_{n-1}} = (-1)^{n-1} a_1,$$

$$a_n \cdot c_1 c_2 \ldots c_n = (-1)^n a_n,$$

Vieta's formulas are usually introduced in school only for the case of real coefficients and roots (often restricted to $n = 2$).

Let $c_1, c_2, \dots, c_n$ be the distinct roots of the polynomial A. Then the factorization of $A(x)$ can be written as

$$A(x) = a_n(x - c_1)^{r_1}(x - c_2)^{r_2} \dots (x - c_n)^{r_k},$$

where r_j denotes the multiplicity of the root c_j for $j = 1, 2, \dots, k$. In this case, it is clear that

$$r_1 + r_2 + \cdots r_k = n = \deg A.$$

REDUCIBILITY OF POLYNOMIALS OVER THE FIELD $\mathbb{R}$. In the ring $\mathbb{R}[x]$, there exist polynomials of degree greater than 1 that are irreducible over $\mathbb{R}$; for example, $x^2 + 1$. However, it turns out that no such irreducible polynomials of degree greater than 2 can exist.

Lemma 5. If $A(x) \in \mathbb{R}[x]$, $c \in \mathbb{C}$ *and* $A(c) = 0$, *then* $A(\bar{c}) = 0$. *In other words, if a complex number* $c = a + bi$ *is a root of a polynomial with real coefficients, then its conjugate* $\bar{c} = a - bi$, *with* $a, b \in \mathbb{R}$, *is also a root.*

Proof. Let

$$A(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

where $a_n, a_{n-1}, \dots, a_1, a_0 \in \mathbb{R}$. Suppose that

$$A(c) = a_n c^n + a_{n-1}c^{n-1} + \cdots + a_1 c + a_0 = 0,$$

for some $c \in \mathbb{C}$. Using the basic properties of complex conjugation, namely

$$\overline{c_1 c_2} = \bar{c_1}\bar{c_2}, \overline{c_1 + c_2} = \bar{c_1} + \bar{c_2},$$

and the fact that $\bar{a} = a$ for all $a \in \mathbb{R}$ (in particular, $\bar{0} = 0$), we obtain

$$\overline{A(c)} = \overline{a_n c^n + a_{n-1}c^{n-1} + \cdots + a_1 c + a_0} = a_n(\bar{c})^n + a_{n-1}(\bar{c})^{n-1} + \cdots + a_1\bar{c} + a_0 = A(\bar{c}).$$

Since $A(c) = 0$, it follows that $\overline{A(c)} = 0$. Thus, $A(\bar{c}) = 0$, which proves the claim. •

Theorem 23. *If* $A(x) \in \mathbb{R}[x]$ *and* $\deg A(x) > 2$, *then the polynomial* $A(x)$ *is reducible over* $\mathbb{R}$.

Proof. Since $\mathbb{R} \subset \mathbb{C}$, the polynomial $A(x)$, with real coefficients, also belongs to $\mathbb{C}[x]$. By the Fundamental Theorem of Algebra, $A(x)$ has at least one root $c = a + bi \in \mathbb{C}$. If $c \in \mathbb{R}$ (i.e., $b = 0$), then $x - c \in \mathbb{R}[x]$ divides $A(x)$. Hence, $A(x)$ is reducible over $\mathbb{R}$. If $c \notin \mathbb{R}$ (i. e., $b \neq 0$), then, by Lemma 5, $\bar{c} = a - bi$ is also a root. Thus, $x - c$ and $x - \bar{c}$ both divide $A(x)$. Since they are distinct irreducible monic polynomials in $\mathbb{C}[x]$, they are coprime by Lemma 4. Therefore, their product $(x - c)(x - \bar{c}) = x^2 - (c + \bar{c})x + c\bar{c}$ divides $A(x)$. Note that $c + \bar{c} = 2a \in \mathbb{R}$ and $c\bar{c} = a^2 + b^2 \in \mathbb{R}$, so this quadratic factor lies in $\mathbb{R}[x]$. Since $\deg A(x) > 2$, the existence of a quadratic divisor in $\mathbb{R}[x]$ implies that $A(x)$ is reducible over $\mathbb{R}$. •

Corollary 6. *Every polynomial of odd degree with real coefficients has at least one real root*.

Proof. Each non-real complex root of a polynomial with real coefficients occurs in a conjugate pair (by Lemma 5). Hence, the linear factors $(x - c)$ and $(x - \bar{c})$ can be combined into quadratic factors with real coefficients. Since the polynomial has odd degree, after pairing all non-real roots there must remain at least one unpaired linear factor. This factor has necessarily real coefficients since together with the quadratic factors it reproduces the original polynomial. Therefore, the unpaired linear factor corresponds to a real root. •

Remark 25. Corollary 6 remains valid in the presence of multiple roots. Indeed, if $c = a + bi$ with $b \neq 0$ is a root of multiplicity r, then $\bar{c} = a - bi$ is also a root of multiplicity r (prove this).

Remark 26. The existence of a real root for a polynomial of odd degree over $\mathbb{R}$ can be proved analytically. The function $y = A(x)$ is defined and continuous on the entire real line $(-\infty, +\infty)$. As $|x| \to +\infty$, the leading coefficient of $A(x)$ grows faster in absolute value than the sum of the remaining terms. Hence, there exists a sufficiently large positive number L such that the sign of $A(L)$ coincides with that of the leading coefficient. Since $\deg A$ is odd, the signs of $A(L)$ and $A(-L)$ are opposite. By the *Bolzano Theorem*, a continuous function on $[-L, L]$ that takes opposite signs at the endpoints, must vanish at least once in that interval. Therefore, $A(x)$ has at least one real root.

Example 26. This example illustrates Theorem 23. At first glance, the polynomial $x^4 + 4$ may appear irreducible over $\mathbb{R}$. However, Theorem 23 implies that this is not the case. One possible approach is to determine all fourth roots of -4 and then combine the conjugate pairs of linear factors (we leave this as an exercise). A simpler method is to use the following algebraic manipulation:

$$x^4 + 4 = (x^4 + 4x^2 + 4) - 4x^2 = (x^2 + 2)^2 - (2x)^2 = [(x^2 + 2) - 2x][(x^2 + 2) + 2x] =$$
$$= (x^2 - 2x + 2)(x^2 + 2x + 2).$$

Thus, $x^4 + 4$ factors into a product of two irreducible quadratics over $\mathbb{R}$.

Theorem 24. (*Factorization of Polynomials over* $\mathbb{R}$) *Any polynomial* $A(x) \in \mathbb{R}[x]$ *can be uniquely factored into irreducible polynomials over* $\mathbb{R}$ *in the following form*:

$$A(x) = a_n(x - c_1)^{r_1} \dots (x - c_k)^{r_k}(x^2 + p_1 x + q_1)^{s_1} \dots (x^2 + p_l x + q_l)^{s_l},$$

where

- $a_n, c_1, \dots, c_k, p_1, q_1, \dots, p_l, q_l \in \mathbb{R}$,
- $r_1, \dots, r_k, s_1, \dots, s_l \in \mathbb{N}$,
- *the degree condition holds*:

$$r_1 + \cdots + r_k + 2s_1 + \cdots + 2s_l = n = \deg A$$

- *and for each quadratic factor, the discriminant is negative*

$p_i^2 - 4q_i < 0$ *for all* $i = 1, \dots, l$. •

Remark 27. The discriminants must be negative because otherwise a quadratic trinomial would factor further into real linear factors, contradicting irreducibility over $\mathbb{R}[x]$.

POLYNOMIALS OVER THE FIELD $\mathbb{Q}$. For polynomials with rational coefficients, the problem of reducibility over the field $\mathbb{Q}$ is more subtle than in the cases of $\mathbb{C}$ and $\mathbb{R}$. Unlike those fields, there is no general universal criterion for irreducibility over $\mathbb{Q}$. One useful tool is the *Eisenstein criterion*, studied in standard algebra textbooks (e.g., D.G. Kurosh, *Course of Higher Algebra*). Despite its name, it is not a full criterion, but rather a *sufficient condition for irreducibility*. For example, it can be applied to show that the polynomial $x^n + x^{n-1} + \cdots + x + 1$ is irreducible over $\mathbb{Q}$ for any $n \geq 0$. This already demonstrates the existence of irreducible polynomials of arbitrarily large degree in $\mathbb{Q}[x]$. Thus, in practice, the main question shifts from general reducibility to the more concrete problem of finding all rational roots of a given polynomial.

It is convenient to regard the coefficients of the polynomial $A \in \mathbb{Q}[x]$ as integers. This does not affect the generality of the problem of finding roots because the roots clearly remain unchanged if the entire polynomial is multiplied by any nonzero number, in particular, by the least common denominator of all fractional coefficients. Thus, we may assume that

$$A(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}(x)$$

is a given polynomial with integer coefficients (i.e., defined over the ring $\mathbb{Z}$, and hence over the field $\mathbb{Q}$). We first consider integer roots, noting that the root 0 (together with its multiplicity) can be identified immediately.

Theorem 25. *If $x_0 \neq 0$ is an integer root of a polynomial $A \in \mathbb{Z}(x)$, then x_0 must be a divisor of the constant term a_0 of A.*

Proof. From

$$A(x_0) = a_n {x_0}^n + a_{n-1}{x_0}^{n-1} + \cdots + a_1 x_0 + a_0 = 0,$$

it follows that

$$a_0 = x_0 \cdot (-a_n {x_0}^{n-1} - a_{n-1}{x_0}^{n-2} - \cdots - a_1).$$

Thus, x_0 divides a_0 ($x_0 \mid a_0$), as claimed. •

By Theorem 25, all integer roots of a polynomial can be determined (or their absence established).

Example 27. Consider the polynomial

$$A(x) = 2x^7 - 2x^6 - 9x^5 - 7x^4 - 4x^3 - 4x^2 - 5x - 3.$$

The divisors of the constant term are ± 1 and ± 3. Hence, any integer root must belong to this set. To test these candidates, it is convenient to use the *Repeated Horner's Scheme* (also referred to as the *Multi-stage* or *Multi-story Method*). In this scheme, the steps ("stories") that produce a nonzero remainder R are omitted (we shade them in practice; see Table 6). Each root, since it may occur with multiplicity, requires at least two successive stories.

Table 6

	2	-2	-9	-7	-4	-4	-5	-3	
1	2	0	-9	-16	-20	-24	-29	-32	1 is not a root.
-1	2	-4	-5	-2	-2	-2	-3	0	
-1	2	-6	1	-3	1	-3	0		
-1	2	-8	9	-12	13	-16			-1 is a root of multiplicity 2.
3	2	0	1	0	1	0			
3	2	6	19	57	172				3 is a root of multiplicity 1.
-3	2	-6	19	-57	172				-3 is not a root.

Thus, the polynomial A has two integer roots: $x = -1$ (with multiplicity 2), and $x = 3$ (with multiplicity 1). The quotient obtained by dividing $A(x)$ by $(x+1)^2(x-3)$ is the fourth-degree polynomial $2x^4 + x^2 + 1$, whose coefficients appear in the last unshaded row of Table 6. This polynomial does not have any real roots since its values are strictly positive for all $x \in \mathbb{R}$.

In Example 27, only four numbers were subject to verification. However, if the constant term has a large number of divisors, some of them can be eliminated in advance, even before applying Horner's scheme, using the following considerations:

1. The direct computation of $A(1)$ and $A(-1)$ does not involve any multiplication and is computationally simple. Since both ± 1 divide any constant term $a_0 \neq 0$, they must always be checked.

2. Any other divisor $x_0 \neq \pm 1$ cannot be a root of A if at least one of the numbers $A(1)/(x_0 - 1)$ or $A(-1)/(x_0 + 1)$ is not an integer. Indeed, for any integer root $x_0 \neq \pm 1$ of $A(x) = (x - x_0)Q(x)$, $Q(x) \in \mathbb{Z}(x)$, it follows that $A(1)/(1 - x_0) = Q(1)$, $A(-1)/(1 + x_0) = Q(-1)$ must both be integers.

Example 28. Consider the polynomial

$$A(x) = 3x^6 - x^5 + 9x^4 + 3x^3 - 56x^2 + 54x - 12.$$

Since $A(1) = 0$, we conclude that $x = 1$ is a root. Using Horner's scheme (or direct polynomial division), we establish that its multiplicity is equal to two. Indeed, dividing $A(x)$ by $(x-1)^2$ gives the polynomial

$$Q_1(x) = 3x^4 + 5x^3 + 16x^2 + 30x - 12,$$

for which, clearly, $x = 1$ is not longer a root. In this case, $Q_1(1) = 42$, $Q_1(-1) = -28$. Hence, -1 is not a root either (otherwise we would divide $Q_1(1)$ by $x + 1$ as many times as the multiplicity of that root). Now, the divisors of the constant term -12 to be tested are: $\pm 2, \pm 3, \pm 4, \pm 6, \pm 12$.

Candidates $x_0 = -3, -4, 6, 12, -12$ are rejected since for them the numbers $42/(x_0 - 1)$ are not integers. Similarly, for $x_0 = 2, 4, -6$, the numbers $-28/(x_0 + 1)$ are not integers. (Candidates already excluded need not be tested again). Thus, only $x_0 = -2$ and $x_0 = 3$ remain, which must be tested using Horner's scheme (Table 7.)

Table 7

	3	5	16	30	-12
-2	3	-1	18	-6	0
-2	3	-7	32	-70	
3	3	8	42	120	

-2 is a root of multiplicity 1.

3 is not a root.

Thus, $A(x)$ has the following integer roots:

$$x_1 = 1 \text{ (of multiplicity 2)}, x_2 = -2 \text{ (of multiplicity 1)}.$$

After dividing $A(x)$ by $(x-1)^2(x+2)$, we get the polynomial

$$Q_2(x) = 3x^3 - x^2 + 18x - 6,$$

which has no integer roots.

To find all rational roots of a polynomial, we use the following result:

Theorem 26. *If $A(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \in \mathbb{Z}(x)$ is a monic polynomial with integer coefficients, then all its rational roots are necessarily integers*.

Proof. Assume the contrary: Suppose $A(p/q) = 0$, where $p \neq 0, q \neq 0, \pm 1$ are integers, and the fraction p/q is irreducible, i.e., $\gcd(p, q) = 1$. Then

$$(p/q)^n + a_{n-1}(p/q)^{n-1} + \cdots + a_1(p/q) + a_0 = 0.$$

Multiplying through by q^{n-1} and rearranging gives

$$\frac{p^n}{q} = -(a_{n-1}p^{n-1} + a_{n-2}p^{n-2}q + \cdots + a_1pq^{n-2} + a_0q^{n-1}).$$

The right-hand side is an integer, so the left-hand side must also be an integer. But since $\gcd(p, q) = 1$, it follows that $\gcd(p^n, q) = 1$. Thus, $\frac{p^n}{q}$ is already in lowest terms, which forces $|q| = 1$. This contradicts the assumption that $q \neq \pm 1$. Therefore, every rational root of a monic polynomial with integer coefficients is necessarily an integer. •

By Theorem 26, all rational roots of the polynomial

$$A(x) = a_nx^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1x + a_0 \in \mathbb{Z}(x),$$

where $a_n \neq 0$, can be determined as follows. Multiply the polynomial by a_n^{n-1} (which does not change the set of roots), and introduce a new formal variable $y = a_nx$. Then the polynomial $a_n^{n-1}A(x)$ can be rewritten as a monic polynomial in y:

$$B(y) = y^n + a_{n-1}y^{n-1} + a_{n-2}a_ny^{n-2} + \cdots + a_1a_n^{n-2}y + a_0a_n^{n-1}.$$

Note that, according with the general definition of the ring K^∞ of sequences and its polynomial subrings, this substitution of the formal variable amounts to expressing the polynomial $(a_0, a_1, \ldots, a_n)$ in terms of the powers of the element $y = (0, a_n, 0, \ldots, 0)$, rather than in terms of $x = (0, 1, 0, \ldots, 0)$. The ring $K[y]$ of polynomials in the variable y is isomorphic to the ring $K[x]$; the natural isomorphism should be written explicitly.

And so, each rational root x_0 of $A(x)$ corresponds uniquely to a rational root $y_0 = a_nx_0$ of $B(y)$. By Theorem 26, this root y_0 must be an integer. Consequently, all rational roots of the original polynomial $A(x)$ can be obtained by finding all integer roots of $B(y)$ and dividing them by a_n.

Thus, in Example 28, the polynomial

$$Q_2(x) = 3x^3 - x^2 + 18x - 6$$

was considered. This polynomial has no integer roots; however, since it is not monic, fractional roots may exist. Multiplying Q_2 by 3^2 and making the substitution $y = 3x$, we obtain the monic polynomial

$$B(y) = y^3 - y^2 + 54y - 54,$$

which, in this case, factors immediately as

$$B(y) = (y - 1)(y^2 + 54).$$

This factorization spares us from testing the numerous divisors of 54. The quadratic factor $y^2 + 54$ is strictly positive for all $y \in \mathbb{R}$, while the linear factor gives the simple root $y_0 = 1$. From this, we deduce the rational root $x_0 = 1/3$ also of multiplicity one (this should be verified rigorously) of the original polynomial

$$A(x) = 3x^6 - x^5 + 9x^4 + 3x^3 - 56x^2 + 54x - 12.$$

Thus, besides the rational roots $x = \frac{1}{3}, 1$, and -2 (the latter two found previously), the polynomial $A(x)$ has no other rational roots.

EMBEDDING A RING INTO A FIELD

As is well known, in the ring $\mathbb{Z}$ of integers, division is not always feasible: not for any $m, n \in \mathbb{Z}$ with $n \neq 0$, there exists $a \in \mathbb{Z}$ such that $m = an$. The need to divide numbers into equal parts – arising from both practical and theoretical considerations – led to the introduction of fractions. This ultimately resulted in the extension of the ring $\mathbb{Z}$ into the field $\mathbb{Q}$ of rational numbers. Let's now generalize and rigorously justify this historical process.

We have already studied field extensions; here, however, we must extend not a field but a ring, in such a way that the new structure strictly contains the original ring (or an isomorphic copy of it) as a subring, while itself

forming a field. Since every field is also a ring, the notion of a *subring* is meaningful (similarly, a *subfield* of a ring is simply a subring that is, in fact, a field). The construction of a field containing a given ring – or an isomorphic image of it – as a subring is called an *embedding of the ring into a field*.

Now, since every field is associative, commutative and free of zero divisors, it follows that each of its subrings also satisfies these properties. Thus, these are necessary conditions for a ring to be embeddable in a field. Remarkably, these conditions are not only necessary but also sufficient: whenever a ring satisfies them, there exists a unique (up to isomorphism) minimal field into which the ring can be embedded.

Theorem 27. (*Field of Fractions and its Uniqueness*) *Any associative and commutative ring K without zero divisors* (*called integral ring*) *admits an embedding into a field, and all minimal fields containing K as a subring are isomorphic to one another*.

Proof. First, consider the trivial ring $(\{0\}, +, \cdot)$. This ring is a subring of the field $\mathbb{Z}_2$, which is minimal since every field contains at least two distinct elements, 0 and 1. Moreover, a field consisting solely of 0 and 1 is unique up to isomorphism (to be proved separately).

Now let $K = (M, +, \cdot)$ be a nontrivial associative and commutative ring without zero divisors. In such a ring, the product of finitely many elements is well defined (associativity), and any equality may be reduced by cancelling a common nonzero factor (absence of zero divisors).

(*Construction of the formal fractions*) Let's define $M_0^2 \subseteq M^2$ so that

$$M_0^2 = \{(x, v) \in M \times M : v \neq 0\}.$$

Since K is nontrivial, we have $M_0^2 \neq \emptyset$. For convenience, we denote an ordered pair (x, v) by the formal fraction x/v. On M_0^2, introduce the relation $(x, v) \sim (y, w) \Leftrightarrow xw = yv$. This is an equivalence relation:

- *Reflexive*: $xv = xv$.
- *Symmetric*: If $xw = yv$, then $yv = xw$.
- *Transitive*: If $xw = yv$ and $yu = zw$, then multiplying gives $xwu = yvu = zwv$, hence $xu = zv$.

The set of equivalence classes under this relation will be denoted by $\widehat{M} = M_0^2 / \sim$. Each class is represented by a formal fraction x/v, corresponding to the equivalence class of (x, v). We now define addition and multiplication on $\widehat{M}$ so as to reproduce the familiar arithmetic operations with fractions.

(*Arithmetic of fractions*) Let $\hat{x}, \hat{y} \in \widehat{M}$ and choose representatives $x/u, y/v$ from these classes. Define

$$\hat{x} + \hat{y} = [(xv + uy)/uv], \ \hat{x} \cdot \hat{y} = [xy/uv],$$

where brackets denote the equivalence class in $\widehat{M}$. Since $u \neq 0, v \neq 0$ and K has no zero divisors, we have $u \cdot v \neq 0$. Hence, both operations are well defined on M_0^2. The independence of representatives holds: if $x'/u' = x/u$ and $y'/v' = y/v$, then $x'u = u'x$ and $y'v = v'y$. It follows that

$$(x'v' + u'y')uv = x'v'uv + u'y'uv = x'u \cdot v'v + u'u \cdot y'v = u'x \cdot v'v + u'u \cdot v'y = u'v'(xv + uy),$$

i.e., $(x'v' + u'y')/u'v' = (xv + uy)/uv$. The analogous verification for multiplication is immediate. Thus, the binary operations defined above are consistent, and $(\widehat{M}, +)$, $(\widehat{M}, \cdot)$ are well defined groupoids.

Both groupoids are associative and commutative. Indeed, let $\hat{x}, \hat{y}, \hat{z} \in \widehat{M}$ and let $x/u, y/v, z/w$ be representatives of these classes. Since $(xv + uy)/uv \in \hat{x} + \hat{y}$, we obtain

$$[(xv + uy)w + uv \cdot z]/(uv \cdot w) = [(xv + uy)/uv] + z/w \in (\hat{x} + \hat{y}) + \hat{z}.$$

On the other hand, since $(yw + vz)/vw \in \hat{y} + \hat{z}$, it follows that

$$[x \cdot vw + u(yw + vz)]/(u \cdot vw) = x/u + [(yw + vz)/vw] \in \hat{x} + (\hat{y} + \hat{z}).$$

But the left-hand side of these two fractions are equal (in fact, in this case they even coincide); hence they represent the same class. Therefore, $(\hat{x} + \hat{y}) + \hat{z} = \hat{x} + (\hat{y} + \hat{z})$. Similarly, the relations

$$\hat{x} + \hat{y} = \hat{y} + \hat{x}, \ (\hat{x}\hat{y})\hat{z} = \hat{x}(\hat{y}\hat{z}), \ \hat{x}\hat{y} = \hat{y}\hat{x}$$

are verified without difficulty.

In the groupoid $(\widehat{M}, +)$, the neutral element 0 is the class of fractions of the form $0/t$ $(t \neq 0)$, and the opposite for the class $\hat{x}$, which contains the fraction x/u, is the class $(-\hat{x})$, which contains $(-x)/u$. Thus, $(\widehat{M}, +)$ is an abelian group. The neutral element $\hat{1}$ of the groupoid $(\widehat{M}, \cdot)$ is the class of fractions of the form t/t $(t \neq 0)$, and the inverse of the class $\hat{x} \neq 0$, containing the fraction x/u $(x, u \neq 0)$, is the class $\hat{x}^{-1}$, which contains u/x. Everything stated in this paragraph, including distributivity $\hat{x}(\hat{y} + \hat{z}) = \hat{x}\hat{y} + \hat{x}\hat{z}$, can be verified directly.

Thus, for the algebraic system denoted by $\hat{P} = (\widehat{M}, +, \cdot)$, all conditions of the field definition have been satisfied (albeit in a different order). Hence, $\hat{P}$ is a field.

(*Embedding of K into $\hat{P}$*) Recall that a *homomorphism* is a mapping between two algebraic structures of the same type that *preserves their operations*. A *bijective homomorphism* is called an *isomorphism* since in this case the inverse mapping is also a homomorphism. Now, let's select a subset $\widehat{M}_K \subset \widehat{M}$ consisting of classes of fractions of the form xt/t $(t \neq 0)$, and show that the system $\widehat{K} = (\widehat{M}_K +, \cdot)$, with the same (induced) operations of addition and multiplication as in the field $\hat{P}$, is a subring isomorphic to the original ring K. To establish this, it suffices to verify that $(\widehat{M}_K, +)$ and $(\widehat{M}_K, \cdot)$ are groupoids, and to construct a bijection $f : M \to \widehat{M}_K$ such that f is simultaneously an isomorphism from $(\widehat{M}, +)$ onto $(\widehat{M}_K, +)$ and from $(\widehat{M}, \cdot)$ onto $(\widehat{M}_K, \cdot)$.

If $\hat{x}, \hat{y} \in \widehat{M}$, $xt/t \in \hat{x}$ and $yt/t \in \hat{y}$, then the class $\hat{x} + \hat{y}$ contains the fraction

$$(xt \cdot t' + t \cdot yt')/(t \cdot t') = (x + y) \cdot tt'/tt',$$

and therefore, also belongs to $\widehat{M}_K$. Similarly, the class $\hat{x} \cdot \hat{y}$ contains the fraction

$$(xt \cdot yt')/(t \cdot t') = xy \cdot tt'/tt',$$

whence $\hat{x} \cdot \hat{y} \in \widehat{M}_K$. The class $-\hat{x}$ contains the fraction $(-x)t/t$ and consequently $-\hat{x} \in \widehat{M}_K$. We now define $f(x) = \hat{x}$ for any $x \in M$, where x is the class of fractions of the form xt/t. We will show that the mapping $f: M \to \widehat{M}_K$ is an isomorphism of the rings K and $\widehat{K}$.

Injectivity: From $\hat{x} = \hat{y}$, i.e., $xt/t = yt'/t'$ for any fractions $xt/t \in \hat{x}$ and $yt'/t' \in \hat{y}$, it follows that $xtt' = ytt'$. By cancelling $tt' \neq 0$, we obtain $x = y$. Hence, f is injective.

Surjectivity: Every element of $\widehat{M}_K$ is of the form xt/t for some $x \in M$.

Homomorphism: From the definitions of addition and multiplication in $\widehat{M}$, one checks that

$$f(x + y) = f(x) + f(y) \text{ and } f(xy) = f(x) \cdot f(y).$$

But these properties were already established when proving that $(\widehat{M}_K, +)$ and $(\widehat{M}_K, \cdot)$ form groupoids. Therefore, f is an isomorphism of rings, and $K \cong \widehat{K} \subseteq \hat{P}$.

Thus, the ring $K = (M, +, \cdot)$ is isomorphic to the subring $\widehat{K} = (\widehat{M}_K +, \cdot)$ of the field $\hat{P} = (\widehat{M}, +, \cdot)$.

(*Uniqueness*) Let P be the minimal field containing K as a subring. For any $x, u \in M$ with $u \neq 0$, in P there exists an element xu^{-1} (not necessarily in K); we assign to it the corresponding class, namely, the element of the field $\hat{P}$ representing the fraction x/u. This correspondence between certain elements of P and all of $\hat{P}$ is well defined since the equality $xu^{-1} = yv^{-1}$ is equivalent to $xv = yu$, that is, the equality of the classes containing fractions x/u and y/v. Moreover, the sum and product in P correspond to the sum and product of their images in $\hat{P}$ since $xu^{-1} + yv^{-1} = (xv + uy)(uv)^{-1}$ and $xu^{-1} \cdot yv^{-1} = (xy)(uv)^{-1}$. Consequently, $\hat{P}$ is isomorphic to some subfield of P, and by minimality, P itself must be of this form. Thus, all minimal fields containing the given ring K as a subring are isomorphic to the field $\hat{P}$, and hence to each other.

(*Conclusion*) Every integral ring K embeds into a field $\hat{P}$, called its field of fractions. Any two minimal fields containing K as a subring are isomorphic to $\hat{P}$. •

If K_0 is an arbitrary associative and commutative ring without zero divisors (i.e., an integral ring), then the polynomial ring $K_0[x]$ has the same properties. The minimal field containing $K_0[x]$ as a subring is called the *field of rational functions over* K_0.

Questions.

1. What is the field of rational functions over a trivial ring? Over the finite field $\mathbb{Z}_2$?
2. Can the field of rational functions over a nontrivial ring K_0 (an integral ring) be obtained by first embedding K_0 into a field P_0 and then constructing the field of rational functions over P_0?

We now assume that for the initial ring K_0 we have taken a field P. It is natural to view the field of rational functions over P as its simple transcendental extension, denoted by $P(x)$, since the element x does not satisfy any algebraic equation over P (why?). The elements of $P(x)$ are fractions of the form $A(x)/B(x)$, $A(x), B(x) \in P[x]$, $B(x) \neq 0$, considered up to cancelation of common factors in the numerator and denominator. Note the distinction between $P(x)$ and $P[x]$. The established expression "rational function" for such fractions is somewhat misleading, because two equal fractions (i.e., the same element of the field $P(x)$) do not necessarily coincide as functions of the variable $x \in P$. For example,

$$x(x-1)/[(x^3+1)(x-1)] = x(x+2) / [(x^3+1)(x+2)]$$

in the sense of equality of fractions over the field $\mathbb{R}$ since

$$x(x-1) \cdot (x^3+1)(x+2) = (x^3+1)(x-1) \cdot x(x+2).$$

However, when consider as functions, they differ: at $x = 1$, the left-hand side is undefined, while the right-hand side takes the value ½; at $x = -2$ the left-hand side equals $2/7$, while the right-hand side is undefined. Such discrepancies occur only at finitely many values of x. After complete reduction of fractions, the resulting functions coincide wherever they are defined (or fail to exist at the same points). Hence, in mathematical analysis, the distinction between "rational function" and "fraction" is not particularly problematic. To simplify notation, we shall omit the explicit "(x)" in the symbols (in capital letters) of polynomials whenever the same variable is understood, but not a specific value.

Let $B \in P[x]$ be a polynomial of positive degree, irreducible over the field P. A fraction of the form A/B^r, $A \in P[x]$, $\deg A < \deg B$, $r \in \mathbb{N}$, is called a *simple fraction with respect to* B. By convention, we also include the zero element of the field $P(x)$ among such fractions, which may be regarded as the special case $\deg A < 0$.

Theorem 28. (*Partial Fractions over a Field*) *Let $C/S \in P(x)$, where $S = aB_1^{r_1}B_2^{r_2} \dots B_k^{r_k}$ is the factorization of the polynomial S into irreducible monic polynomials $B_1, B_2, \dots, B_k$ with exponents $r_1, r_2, \dots, r_k$ and $a \in P$. Then the fraction C/S can be expressed, in one and only one way, as a sum of a polynomial and simple fractions with respect to $B_1, B_2, \dots, B_k$, whose denominators are among*

$$B_1, B_1^2, \dots, B_1^{r_1}, B_2, B_2^2, \dots, B_2^{r_2}, \dots, B_k, B_k^2, \dots, B_k^{r_k}.$$

Proof.

(*Reduction of the scalar factor*) Because $a \neq 0$ (otherwise $S = 0$, which is impossible), we may assume without loss of generality that $a = 1$. Indeed, if $a \neq 1$, we replace the numerator C by $a^{-1}C$.

(*Existence*) We first prove the existence of the desired representation, beginning with the special case $k = 1$, i.e., when the denominator of the given fraction has the form $S = B^r$.

Dividing C by B, we obtain $C = BQ_{r-1} + A_r$, $\deg A_r < \deg B$, where the notation of the quotient Q_{r-1} and remainder A_r will soon prove useful. Hence,

$$C/S = C/B^r = Q_{r-1}/B^{r-1} + A_r/B^r,$$

where the second term is a simple fraction (possibly zero). Dividing Q_{r-1} further by B, we obtain $Q_{r-1} = BQ_{r-2} + A_{r-1}$, $\deg A_{r-1} < \deg B$, whence

$$Q_{r-1}/B^{r-1} = Q_{r-2}/B^{r-2} + A_{r-1}/B^{r-1}.$$

Therefore,

$$C/S = Q_{r-2}/B^{r-2} + A_{r-1}/B^{r-1} + A_r/B^r,$$

where the last two terms are simple fractions. Continuing this process, we arrive at

$$C/S = Q_1/B + A_2/B^2 + \cdots + A_{r-1}/B^{r-1} + A_r/B^r.$$

Finally, dividing Q_1 by B, we obtain

$$C/S = Q + A_1/B + A_2/B^2 + \cdots + A_r/B^r,$$

where $Q, A_1, A_2, \ldots, A_r \in P[x]$, and all fractions on the right-hand side are simple. This proves existence for $k = 1$.

(*Induction on* k) Let $k > 1$. Recall that our convention regarding the omission of the multiplication symbol sometimes makes it possible to write products without brackets. In the expansion

$$S = B_1^{r_1} \cdot B_2^{r_2} \ldots B_k^{r_k}$$

the two factors $B_1^{r_1}$ and $B_2^{r_2} \ldots B_k^{r_k}$ are relatively prime. The reason is that the B_i are distinct irreducible, $\gcd\left(B_1, B_2^{r_2} \ldots B_k^{r_k}\right) = 1$, hence also $\gcd\left(B_1^{r_1}, B_2^{r_2} \ldots B_k^{r_k}\right) = 1$. By Corollary 4, there exist polynomials $U, V \in P[x]$ such that

$$U \cdot B_1^{r_1} + V \cdot B_2^{r_2} \ldots B_k^{r_k} = 1.$$

It follows that $C = CV \cdot B_2^{r_2} \ldots B_k^{r_k} + CU \cdot B_1^{r_1}$, and hence

$$C/S = CV/B_1^{r_1} + CU/\left(B_2^{r_2} \ldots B_k^{r_k}\right).$$

The first term, as we have already established, admits a representation as the sum of a polynomial and simple fractions with respect to B_1. Proceeding with the second term in the same way as for the case $k = 2$, and analogously treating the original fraction C/S when $k > 2$, after k steps – and collecting all polynomials of type Q – we arrive at the desired representation (the general form of which will be introduced a little later). This completes existence by induction.

(*Uniqueness*) Suppose that

$$C/S = Q + \cdots + A/B_i^r + \cdots = \tilde{Q} + \cdots + \tilde{A}/B_i^r + \cdots$$

are two different representations, where $A \neq \tilde{A}$ $(1 \leq i \leq k, 1 \leq r \leq r_i)$. Thus, there exist two simple fractions with the same denominator B_i^r, but different numerators. Such fractions must necessarily occur (why?). Among all such pairs, let's choose one with the largest exponent r. Multiplying both sides of

$$\left(A - \tilde{A}\right)/B_i^r = \tilde{Q} - Q + \cdots$$

by

$$B_1^{r_1} \ldots B_{i-1}^{r_{i-1}} B_i^{r-1} B_{i+1}^{r_{i+1}} \ldots B_k^{r_k},$$

we obtain an equality in which the left-hand side still contains simple fractions, while the right-hand side becomes a nonzero polynomial – a contradiction (why?).

Hence, uniqueness holds, and the theorem is proven. •

Remark 28. To represent the original fraction as the sum of a polynomial and simple fractions in general form, note that the numerators of the terms associated with different denominators B_i are distinct. Therefore, their notation requires two indices: A_{ij} denotes the numerator corresponding to the denominator B_i^j for $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, r_i$. The representation, whose existence and uniqueness are guaranteed by Theorem 28, takes the general form

$$C/S = Q + A_{11}/B_1 + A_{12}/B_1^2 + \cdots + A_{1r_1}/B_1^{r_1} +$$
$$+ A_{21}/B_2 + A_{22}/B_2^2 + \cdots + A_{2r_2}/B_2^{r_2} + (8)$$
$$\ldots\ldots\ldots\ldots$$
$$+ A_{k1}/B_k + A_{k2}/B_k^2 + \cdots + A_{kr_k}/B_k^{r_k},$$

the indices are read digit by digit: e.g., A_{11} is read as "A one one" (not "A eleven"), and A_{1r_1} as "A one r_1". By convention, the indices are not separated by a comma (sometimes one writes $A_{(1,\,r_1)}$, but not $A_{1,1}$). In (8), $Q, A_{ij} \in P[x]$, $\deg A_{ij} < \deg B_i$ for all $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, r_i$. If $A_{ij} = 0$, then the corresponding fraction represents the zero element of the field $P(x)$ and that term disappears from (8) when its numerator vanishes.

In the particular case $P = \mathbb{R}$, considering the reducibility of polynomials over the real field, we obtain the following result.

Corollary 7. *Every fraction* $C/S \in P(x)$, *where*

$$S = a(x - c_1)^{r_1} \ldots (x - c_k)^{r_k}(x^2 + p_1 x + q_1)^{s_1} \ldots (x^2 + p_l x + q_l)^{s_l},$$

with

$a, c_1, \ldots, c_k, p_1, \ldots, q_l \in \mathbb{R}$, $r_1, \ldots, r_k, s_1, \ldots, s_l \in \mathbb{N}$, $p_i^2 - 4q_i < 0$ *for all* $i = 1, \ldots, l$,

is uniquely representable as a sum of a polynomial over $\mathbb{R}$ *and simple fractions*:

$$C/S = Q + \alpha_{11}/(x - c_1) + \alpha_{12}/(x - c_1)^2 + \cdots + \alpha_{1r_1}/(x - c_1)^{r_1} +$$
$$+ \cdots + \alpha_{k1}/(x - c_k) + \alpha_{k2}/(x - c_k)^2 + \cdots + \alpha_{kr_k}/(x - c_k)^{r_k} +$$
$$+ (\beta_{11}x + \gamma_{11})/(x^2 + p_1 x + q_1) + (\beta_{12}x + \gamma_{12})/(x^2 + p_1 x + q_1)^2 +$$
$$+ \cdots + \left(\beta_{1s_1}x + \gamma_{1s_1}\right)/(x^2 + p_1 x + q_1)^{s_1} + \cdots +$$

$$+ (\beta_{l1}x + \gamma_{l1})/(x^2 + p_l x + q_l) + (\beta_{l2}x + \gamma_{l2})/(x^2 + p_l x + q_l)^2 + \\ + \cdots + \left(\beta_{ls_l}x + \gamma_{ls_l}\right)/(x^2 + p_l x + q_l)^{s_l},$$

where all coefficients

$\alpha_{11}, \dots, \alpha_{kr_k}, \beta_{11}, \dots, \beta_{ls_l}, \gamma_{11}, \dots, \gamma_{ls_l} \in \mathbb{R}$. •

In practice, it is convenient to find such a representation not by the general scheme of Theorem 28, but rather by the *method of undetermined coefficients*. To avoid cumbersome formulas, we illustrate the problem with an example.

Example 29. (*Partial fraction decomposition*) Consider

$$C = 8x^7 + 4x^6 + 2x^5 + 11x^4 + 15x^3 + 9x^2 + 10x + 11,$$

$$S = 8x^6 + 4x^5 - 2x^4 - 5x^3 - 2x^2 + 4x - 1.$$

(*Division of C by S*)

$$\frac{C}{S} = x + \frac{4x^5+16x^4+17x^3+5x^2+11x+11}{8x^6+4x^5+2x^4-5x^3-2x^2+4x-1}.$$

(*Factorization*)

$$= x + \frac{(x+1)\left(4x^4+12x^3+5x^2+11\right)}{(x+1)(2x-1)^3(x^2+x+1)}.$$

(*Cancellation and monic denominator*)

$$= x + \frac{\frac{1}{2}x^4+\frac{3}{2}x^3+\frac{5}{8}x^2+\frac{11}{8}}{\left(x-\frac{1}{2}\right)^3(x^2+x+1)}.$$

Here we not only reduced the fraction by cancelling $(x + 1)$, but also made the denominator monic by multiplying the numerator by $1/8$. In the resulting fraction we have

$$k = l = 1,\ c_k = c_1 = \frac{1}{2},\ r_k = r_1 = 3,\ p_l = p_1 = q_l = q_1 = 1, s_l = s_1 = 1.$$

We now write the fraction as a sum of simple fractions with undetermined coefficients (not yet defined explicitly):

$$\frac{\frac{1}{2}x^4+\frac{3}{2}x^3+\frac{5}{8}x^2+\frac{11}{8}}{\left(x-\frac{1}{2}\right)^3(x^2+x+1)} = \frac{\alpha}{\left(x-\frac{1}{2}\right)} + \frac{\beta}{\left(x-\frac{1}{2}\right)^2} + \frac{\gamma}{\left(x-\frac{1}{2}\right)^3} + \frac{\delta x+\varepsilon}{(x^2+x+1)}.$$

(*Reduction to a common denominator*) Multiplying both sides by $\left(x - \frac{1}{2}\right)^3 (x^2 + x + 1)$, we obtain the polynomial identity

$$\alpha\left(x - \frac{1}{2}\right)^2 (x^2 + x + 1) + \beta\left(x - \frac{1}{2}\right)(x^2 + x + 1) + \gamma(x^2 + x + 1) + (\delta x + \varepsilon)\left(x - \frac{1}{2}\right)^3 \\ = \frac{1}{2}x^4 + \frac{3}{2}x^3 + \frac{5}{8}x^2 + \frac{11}{8}$$

(*Expansion and coefficient comparison*) Expanding the left-hand side, grouping by powers of x, and equating coefficients gives the system

$$\alpha + \delta = 1/2,$$
$$\beta - \frac{3}{2}\delta + \varepsilon = 3/2,$$
$$\frac{1}{4}\alpha + \frac{1}{2}\beta + \gamma + \frac{3}{4}\delta - \frac{3}{2}\varepsilon = 5/8,$$
$$-\frac{3}{4}\alpha + \frac{1}{2}\beta + \gamma - \frac{1}{8}\varepsilon = 0,$$
$$\frac{1}{4}\alpha - \frac{1}{2}\beta + \gamma - \frac{1}{8}\varepsilon = 11/8,$$

(*Solving the system*)

$$\alpha = \frac{3}{2}, \beta = 0, \gamma = 1, \delta = -1, \varepsilon = 0.$$

(*Final representation*)

$$\frac{C}{S} = x + \frac{\frac{3}{2}}{x-\frac{1}{2}} + \frac{1}{\left(x-\frac{1}{2}\right)^3} - \frac{x}{x^2+x+1} = x + \frac{3}{2x-1} + \frac{8}{(2x-1)^3} - \frac{x}{x^2+x+1}.$$

Remarks 29. In general, the system of equations for the unknowns $\alpha_{11}, \dots, \beta_{ls_l}, \gamma_{ls_l}$ will always have a unique solution. This follows automatically from the existence and uniqueness of the partial fraction decomposition (Corollary 7). (As an exercise, one may verify that the number of equations coincides with the number of unknowns, though this fact alone does not guarantee solvability or uniqueness; the general theory of such systems is studied in linear algebra courses.)

A separate justification for the process described in Corollary 7 is unnecessary, as it constitutes an immediate consequence of the proof of Theorem 28 itself. Only one subtle point (which, apparently, not everyone noticed) requires clarification. Specifically, in equality (8) we immediately took the quotient of dividing C by S as Q. Then, in the course of the proof, this polynomial gradually developed from many terms. One might ask: could this process introduce any additional terms beyond the original quotient?

In fact, from the uniqueness of the representation $C = SQ + R$, $\deg R < \deg S$, it follows that the representation $C/S = Q + R/S$ is also unique, where the fraction R/S is *proper* (the degree of the numerator is less than that of the denominator). Since every simple fraction is proper, the only gap in the proof is to confirm that the *sum of*

proper fractions remains a proper fraction. We recommend proving this general statement independently, and, as a simpler exercise, verifying that the product of proper fractions is also a proper fraction.

Question. Will the ratio of two proper fractions itself be a proper fraction?

Finally, in the context of mathematical analysis – where Theorem 28 and Corollary 7 are applied primarily to the integration of rational functions – other methods for decomposing a proper fraction into a sum of simple fractions are also studied.

MATRICES

Consider a nine-story building with five entrances. Let a_{ij} denote the total number of residents in those apartments located on the *i*-th floor whose doors open to the stairwell of the *j*-th entrance. The distribution of residents can then be represented by the matrix

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{91} & a_{92} & a_{93} & a_{94} & a_{95} \end{bmatrix},$$

which has nine rows and five columns. This matrix describes the complete distribution of residents in the building according to stairwell. The entries a_{ij} need not be numerical; they may, for example, represent sets (such as lists of residents).

(*General definition*) A $(p \times q)$-matrix over a set M is a function

$$A: \{1, 2, \dots, p\} \times \{1, 2, \dots, q\} \longrightarrow M,$$

which assigns to each ordered pair (i, j) with $i \in \{1, 2, \dots, p\}$, $j \in \{1, 2, \dots, q\}$ an element $a_{ij} \in M$. The notation for a matrix is written as

$$A = (a_{ij}) = (a_{ij})_{p\times q} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{bmatrix},$$

where the indices p and q, denoting the numbers of rows and columns, may be omitted when the context makes them clear. Two $(p \times q)$-matrices $A = (a_{ij})_{p\times q}$ and $B = (b_{ij})_{p\times q}$ are said to be *equal*, written $A = B$, if and only if $a_{ij} = b_{ij}$ for all $i \in \{1, 2, \dots, p\}$ and $j \in \{1, 2, \dots, q\}$. It is not meaningful to speak of equality or inequality between matrices of different sizes (even if they share the same p or q).

Let $K = (M, +, \cdot)$ be a ring. The sum of two $(p \times q)$-matrices with entries in K (that is, in M) is defined *componentwise*, meaning entry by entry. If $A = (a_{ij})_{p\times q}$, $B = (b_{ij})_{p\times q}$, then

$$A + B = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1q} \\ b_{21} & b_{22} & \cdots & b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pq} \end{bmatrix} =$$

$$= \begin{bmatrix} a_{11}+ b_{11} & a_{12}+ b_{12} & \cdots & a_{1q}+ b_{1q} \\ a_{21}+ b_{21} & a_{22}+ b_{22} & \cdots & a_{2q}+ b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1}+ b_{p1} & a_{p2}+ b_{p2} & \cdots & a_{pq}+ b_{pq} \end{bmatrix} = (a_{ij} + b_{ij})_{p\times q}.$$

With respect to addition, the set of all $(p \times q)$-matrices (with fixed p and q) over K forms an abelian group. The neutral element is the zero $(p \times q)$-matrix, denoted by $(0_{ij})_{p\times q}$, whose entries are all equal to the zero element of the ring K. For any matrix $A = (a_{ij})$, the additive inverse is given by $-A = (-a_{ij})$. Defining the product of matrices is a more subtle matter, but it is also considerably more interesting.

The *componentwise* multiplication of matrices, $(a_{ij})_{p\times q} \cdot (b_{ij})_{p\times q} = \big((a_{ij}) \cdot (b_{ij})\big)_{p\times q}$ gives rise to a ring of $(p \times q)$-matrices that inherits the essential properties of the underlying ring: commutativity, associativity, the existence of a unity, the absence of zero divisors, and the invertibility of nonzero elements. This componentwise product, introduced by J. Hadamard, is well known in combinatorial analysis. However, the corresponding matrix ring, considered as an algebraic system, is of particular interest not because it involves new objects such as matrices, but because its structure can be more naturally explained through the already well-studied framework of polynomials.

Let's suppose that, for infinite sequences of elements of a ring K, we define multiplication in the same componentwise fashion as addition:

$$(a_0, a_1, a_2, \dots) \cdot (b_0, b_1, b_2, \dots) = (a_0b_0, a_1b_1, a_2b_2, \dots).$$

Clearly, this again produces a ring that is structurally as "good" as the original ring K (a fact that can be checked and verify). However, in this construction the situation is quite limited: if an element d is given by an expression such as $ab + c$, then the *i*-th term of the resulting sequence satisfies $d_i = a_ib_i + c_i$, $i = 0, 1, 2, \dots$ so that each term depends only on the corresponding terms of the original sequences. No interaction occurs between different indices. In this sense, the resulting ring of sequences is nothing more than a "mechanical mixture" of infinitely many independent copies of the ring K.

By contrast, consider the alternative definition of multiplication:

$$(a_0, a_1, a_2, \dots) \cdot (b_0, b_1, b_2, \dots) = (a_0 b_0,\ a_0 b_1 + a_1 b_0,\ a_0 b_2 + a_1 b_1 + a_2 b_0, \dots).$$

Here, the *i*-th term of the product (for $i > 0$) is influenced not only by the *i*-th terms of the factors but also by earlier terms. This construction leads precisely to the polynomial ring $K[x]$, which represents a qualitatively new algebraic system. Such a fruitful definition of multiplication is not merely a clever invention, but rather a reflection of deeper quantitative principles of the real world – principles that find natural mathematical expression in the operations on polynomials. As other examples – such as the field $\mathbb{Q}(\sqrt{2})$, the field of complex numbers, and the subsequent historical development of algebra – demonstrate, such "linking of different components" proves fruitful precisely when it is incorporated into the definition of multiplication rather than addition. Naturally, however, statements of this kind should not be interpreted categorically since doing so would only restrict scientific creativity.

In order to obtain a "linked" multiplication of matrices rather than an entry-by-entry product, we consider the following example.

Example 30. Suppose there are 3 products from which 5 different dishes are prepared by mixing them in various proportions. Let a_{ij} denote the number of grams of the *j*-th product required to prepare one gram of the *i*-th dish ($i = 1, 2, 3, 4, 5$; $j = 1, 2, 3$). Furthermore, each gram of the *j*-th product contains b_{j1} grams of proteins, b_{j2} grams of water, b_{j3} grams of fat and b_{j4} grams of carbohydrates (these are referred to as the 1st, 2nd, 3rd, and 4th components of the product, respectively). The task is to determine the component composition of each dish.

To find the amount c_{ik} (in grams) of the *k*-th component in the *i*-th dish ($k = 1, 2, 3, 4$), one simply sums the contributions of the three products. For instance, the fat content ($k = 3$) of the second dish ($i = 2$) is obtained by adding up the quantities contributed by the first, second, and third products: $c_{23} = a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33}$. In general, the entries of the desired matrix

$$C = (c_{ik})_{5\times 4}$$

are obtained from the given matrices

$$A = \left(a_{ij}\right)_{5\times 3},\ B = \left(b_{jk}\right)_{3\times 4},$$

according to the rule

$$c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + a_{i3}b_{3k},\ i = 1, \dots, 5,\ k = 1, \dots, 4. \quad (9)$$

Example 31. Here is a more abstract instance. Let the variables x_1, x_2, x_3, x_4, x_5 be expressed through the variables y_1, y_2, y_3 by the equalities

$$\begin{aligned} x_1 &= a_{11}y_1 + a_{12}y_2 + a_{13}y_3, \\ x_2 &= a_{21}y_1 + a_{22}y_2 + a_{23}y_3, \\ x_3 &= a_{31}y_1 + a_{32}y_2 + a_{33}y_3, \\ x_4 &= a_{41}y_1 + a_{42}y_2 + a_{43}y_3, \\ x_5 &= a_{51}y_1 + a_{52}y_2 + a_{53}y_3, \end{aligned} \quad (10)$$

where a_{ij} are constants (for example, real numbers). The variables y_1, y_2, y_3, in turn, are expressed through z_1, z_2, z_3, z_4 as follows:

$$\begin{aligned} y_1 &= b_{11}z_1 + b_{12}z_2 + b_{13}z_3 + b_{14}z_4, \\ y_2 &= b_{21}z_1 + b_{22}z_2 + b_{23}z_3 + b_{24}z_4, \\ y_3 &= b_{31}z_1 + b_{32}z_2 + b_{33}z_3 + b_{34}z_4. \end{aligned} \quad (11)$$

It is required to express the x-variables directly in terms of the z-variables. The solution is obvious: replace all the y-variables in (10) with their expressions from (11), expand the brackets and group the z-variables. As a result, we obtain

$$\begin{aligned} x_1 &= c_{11}z_1 + c_{12}z_2 + c_{13}z_3 + c_{14}z_4, \\ x_2 &= c_{21}z_1 + c_{22}z_2 + c_{23}z_3 + c_{24}z_4, \\ &\vdots \qquad\qquad \vdots \\ x_5 &= c_{51}z_1 + c_{52}z_2 + c_{53}z_3 + c_{54}z_4, \end{aligned} \quad (12)$$

where the entries of the matrix $C = (c_{ik})$ are expressed in terms of the entries of the matrices $A = \left(a_{ij}\right)$ and $B = \left(b_{jk}\right)$ by the same equalities (9) of Example 30.

Of course, it was not the problem of rational nutrition, but the theory of *linear transformations* that served as the basis for the creation of the theory of matrices, in which precisely this method of obtaining the matrix C from A and B plays a key role.

Let $A = \left(a_{ij}\right)_{p\times q}$ and $B = \left(b_{jk}\right)_{q\times r}$ be matrices over a ring K. The sizes of A and B are chosen so that the number of columns of the first matrix equals the number of rows of the second. The product $A \cdot B$ is defined to be the matrix $C = (c_{ik})_{p\times r}$ with entries

$$\sum_{j=1}^{q} a_{ij}b_{jk}\ (i = 1, 2, \dots, p;\ k = 1, 2, \dots, r).$$

Whenever matrices are multiplied without explicitly indicating their sizes, it is always assumed that their dimensions are compatible as described above; otherwise, the matrix product is not defined. Alternatively, without introducing C explicitly, the product of a $(p \times q)$-matrix A by a $(q \times r)$-matrix B can be written as

$$A \cdot B = (a_{ij})_{p\times q} \cdot (b_{jk})_{q\times r} = \left(\sum_{j=1}^{q} a_{ij}b_{jk}\right)_{p\times r},$$

which is, as expected, a $(p \times r)$-matrix.

Matrix multiplication over a ring K is not commutative, even when the ring K itself is commutative.

First, if $A = (a_{ij})_{p\times q}$ and $B = (b_{ij})_{p\times q}$ with $p \neq q$, then equality $A \cdot B = B \cdot A$ does not even make sense. Indeed:

- The product $A \cdot B$ is only defined when the number of columns of A equals the number of rows of B.
- If A is $p \times q$ and B is $p \times q$ then $A \cdot B$ is undefined unless $q = p$.
- Even if $p = r$, the products

$$(a_{ij})_{p\times q} \cdot (b_{jk})_{q\times p} = (A \cdot B)_{p\times p} \text{ and } (b_{ij})_{q\times p} \cdot (a_{jk})_{p\times q} = (B \cdot A)_{q\times q}$$

are of different sizes ($p \times p$ vs. $q \times q$) and thus cannot be equal.

Second, even when $p = q = r$ so that both products exist and have the same size, they are *not necessarily equal*, as shown by the following example.

Example 32. Let

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

over the ring $\mathbb{Z}$ ($p = q = r = 2$). Then

$$AB = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, BA = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix},$$

which are not equal.

Therefore, matrix multiplication is not commutative. Moreover, while both distributive laws

$$A(B + C) = AB + AC \text{ and } (B + C)A = BA + CA$$

hold, they require proof. We shall prove only the first identity since the argument for the second is similar.

Claim 1. *Let $A = (a_{ij})_{p\times q}$, $B = (b_{jk})_{q\times r}$ and $C = (c_{jk})_{q\times r}$ be matrices over a ring K. Then*

$$A(B + C) = AB + AC$$

whenever the sizes make the corresponding addition and multiplication defined.

Proof. Matrix addition is defined only when the summands have the same dimensions. Thus,

$$B + C = (b_{jk} + c_{jk})_{q\times r}.$$

Then

$$A(B + C) = \left(\sum_{j=1}^{q} a_{ij}(b_{jk}+c_{jk})\right)_{p\times r} = \left(\sum_{j=1}^{q} (a_{ij}b_{jk} + a_{ij}c_{jk})\right)_{p\times r}$$

$$= \left(\sum_{j=1}^{q} a_{ij}b_{jk} + \sum_{j=1}^{q} a_{ij}c_{jk}\right)_{p\times r} = \left(\sum_{j=1}^{q} a_{ij}b_{jk}\right)_{p\times r} + \left(\sum_{j=1}^{q} a_{ij}c_{jk}\right)_{p\times r} = AB + AC.$$

This equality holds because the ring K satisfies the distributive law $a(b + c) = ab + ac$ •

Claim 2. *Let $A = (a_{ij})_{p\times q}$, $B = (b_{jk})_{q\times r}$, $C = (c_{kl})_{r\times s}$ be matrices over a ring K whose multiplication is associative. Then $A \cdot BC = AB \cdot C$ whenever the sizes make the products defined.*

Proof.

$$A \cdot BC = (a_{ij})_{p\times q} \cdot \left(\sum_{k=1}^{r} b_{jk}c_{kl}\right)_{q\times s} = \left(\sum_{j=1}^{q} a_{ij}\left(\sum_{k=1}^{r} b_{jk}c_{kl}\right)\right)_{p\times s}$$

$$= \left(\sum_{j=1}^{q}\sum_{k=1}^{r} a_{ij} \cdot b_{jk}c_{kl}\right)_{p\times s} = \left(\sum_{j=1}^{q}\sum_{k=1}^{r} a_{ij}b_{jk} \cdot c_{kl}\right)_{p\times s} = \left(\sum_{k=1}^{r}\sum_{j=1}^{p} a_{ij}b_{jk} \cdot c_{kl}\right)_{p\times s}$$

$$= \left(\sum_{k=1}^{r}\left(\sum_{j=1}^{p} a_{ij}b_{jk}\right)c_{kl}\right)_{p\times s} = \left(\sum_{j=1}^{p} a_{ij}b_{jk}\right)_{p\times r} \cdot (c_{kl})_{r\times s} = AB \cdot C$$

•

Remark 30. (*Entrywise explanation*) Fix $i \in \{1, \dots, p\}$ and $l \in \{1, \dots, s\}$. Then:

$(A \cdot BC)_{il} = \sum_{j=1}^{q} a_{ij}(BC)_{jl}$ (definition of product)

$= \sum_{j=1}^{q} a_{ij}\left(\sum_{k=1}^{r} b_{jk}c_{kl}\right)$ (definition of $(BC)_{jl}$)

$= \sum_{j=1}^{q}\sum_{k=1}^{r} a_{ij} \cdot b_{jk}c_{kl}$ (expand the double sum)

$= \sum_{j=1}^{q}\sum_{k=1}^{r} a_{ij}b_{jk} \cdot c_{kl}$ (associativity in K)

$= \sum_{k=1}^{r}\sum_{j=1}^{p} a_{ij}b_{jk} \cdot c_{kl}$ (finite sums can be reordered)

$= \sum_{k=1}^{r}\left(\sum_{j=1}^{p} a_{ij}b_{jk}\right)c_{kl}$ (group the inner sum)
$= \sum_{k=1}^{r}(AB)_{ik}\,c_{kl}$ (definition of $(AB)_{ik}$)
$= (AB \cdot C)_{il}$ (definition of product).

Since this holds for all i, l, we conclude $A \cdot BC = AB \cdot C$

(*Key properties used*)

- Associativity of multiplication in K: $a_{ij} \cdot b_{jk}c_{kl} = a_{ij}b_{jk} \cdot c_{kl}$.
- Commutativity and associativity of addition in K: finite sums can be regrouped and reordered.

Due to the associativity of matrix multiplication, the solution to the previous problem of variable transformations (see Example 31) can be written in a very compact form.

Namely, along with the coefficient matrices

$$A = \left(a_{ij}\right)_{5\times3},\ B = \left(b_{jk}\right)_{3\times4},\ C = (c_{kl})_{5\times4},$$

we also introduce the variable column matrices:

$$X = \left(x_j\right)_{5\times1} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix},\ Y = (y_k)_{3\times1} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix},\ Z = (z_l)_{4\times1} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix}.$$

Then the system of equations (10) can be written equivalently as the single matrix equation

$$X = AY. \quad (10A)$$

Similarly, the system (11) is equivalent to the matrix equation

$$Y = BZ. \quad (11A)$$

Substituting the expression for Y from (11A) into (10A) and using the associativity of matrix multiplication, we obtain

$$X = A \cdot BZ = AB \cdot Z = CZ,$$

where $C = AB$.

But the equality $X = CZ$ is precisely the matrix form of the system (12).

Strictly speaking, the entries of these matrices do not belong to the original ring K over which the coefficient matrices A and B are defined, but rather to the polynomial ring

$$K[x_1, \dots, x_5, y_1, y_2, y_3, z_1, \dots, z_4]$$

in twelve variables over K. However, if we interpret these variables as taking values in K, then all of the above is valid without additional justification.

If the ring K has unity 1, then the identity $(p \times p)$-matrix

$$I_p = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{p\times p}$$

serves as the *left neutral element* with respect to multiplication for all $(p \times q)$-matrices, and the identity $(q \times q)$-matrix I_q serves as the *right neutral element*. Indeed, for any $(p \times q)$-matrix A over K, one has

$$I_p \cdot (A)_{p\times q} = (A)_{p\times q} \text{ and } (A)_{p\times q} \cdot I_q = (A)_{p\times q}.$$

However, when $p \neq q$ there is *no two-sided neutral element*, because the set of $(p \times q)$-matrices does not form a groupoid under matrix multiplication. This resolves the apparent contradiction with the remark on the uniqueness of the neutral element in a groupoid (Theorem 1).

Example 33. The following calculations illustrate that the product of two nonzero matrices over K may be the zero matrix even when the ring K itself has no zero divisors:

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = (0)_{2\times2},\ [1 \quad 0 \quad 1] \cdot \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & -1 & 0 \end{bmatrix} = [0 \quad 0 \quad 0] = (0)_{1\times3}.$$

We shall present additional properties of matrix multiplication as needed.

From the discussion above, it is evident that the set $M(K)$ of all matrices over an arbitrary ring K (even over a field) does not form a satisfactory algebraic structure of ring type with respect to matrix addition and multiplication. The same holds for any subset consisting of $(p \times q)$-matrices with fixed p and q whenever $p \neq q$.

Only when $p = q$ does the set $M_p(K)$ of all *square* $(p \times p)$*-matrices* form a ring $\left(M_p(K), +, \cdot\right)$. For brevity, we shall continue to denote this ring simply by $M_p(K)$. The fact that $M_p(K)$ is indeed a ring – and that properties such as associativity and the existence of a multiplication identity (but not necessarily commutativity or the absence of zero divisors) carry over from the base ring K – can be established by collecting and organizing the relevant statements and examples presented earlier.

The ring $M_1(K)$ is clearly isomorphic to K; in fact, it essentially coincides with K if we ignore the notational distinction between the symbols $\left(a_{ij}\right)_{1\times1} = [a_{11}]$ and a_{11}. (You could say this, though it brings to mind O. Henry's remark: "She sat in pink dreams and in a kimono of the same color." And from which novel is this, by the way?)

Before examining the rings $M_p(K)$ in detail, let's first consider several important examples of transformations of $(p \times q)$-matrices achieved by multiplying a given matrix by suitable *transforming matrices*. Throughout, we shall assume that the ring K is associative and has a multiplicative identity 1.

Transformation I. (*Row permutation, also known as row swap*)

Suppose that the matrix $A = (a_{ij})_{p\times q}$ $(p > 1)$ over a field K needs to be rearranged as follows: Place the first row of A in the i_1-th position, the second row in the i_2-th position ,…, and the p-th row in the i_p-th position. This arrangement can be achieved by multiplying A on the left by a suitable $(p \times p)$ permutation matrix

$$S\begin{bmatrix} 1 & 2 & \dots & p \\ i_1 & i_2 & \dots & i_p \end{bmatrix},$$

in which each column contains exactly one entry equals to 1 and $p-1$ entries equal to 0; the 1 in the first column is located in the i_1-th row, the 1 in the second column is in the i_2-th row, and so on. For example:

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} a & b & c & d \\ x & y & z & t \\ \alpha & \beta & \gamma & \delta \end{bmatrix} = \begin{bmatrix} \alpha & \beta & \gamma & \delta \\ a & b & c & d \\ x & y & z & t \end{bmatrix}.$$

where the left factor is the transforming matrix

$$S\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}.$$

Transformation II. (*Row replacement by row sum*)

Consider the following modification of a matrix A: the entries of the i-th row are replaced by the sum of themselves and λ times the corresponding entries of the k-th row, where $\lambda \in K$ is fixed and $i \neq k$. The k-th row itself remains unchanged. In other words, the transformation

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{iq} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kq} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{bmatrix} \to \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1} + \lambda a_{k1} & a_{i2} + \lambda a_{k2} & \cdots & a_{iq} + \lambda a_{kq} \\ \vdots & & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kq} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{bmatrix}$$

is equivalent to left-multiplication of A by the matrix $E_{ik}(\lambda) \in M_p(K)$, which is identical to the identity matrix I_p except that the zero element in the (i,k)-position is replaced by λ. Similarly, right-multiplication of A by the matrix $E_{kj}(\lambda) \in M_q(K)$ (with $k \neq j$) means that the entries of the j-th column are replaced by the sum of themselves and λ times the corresponding entries of the k-th column. For example:

$$\begin{bmatrix} a & b & c & d \\ x & y & z & t \\ \alpha & \beta & \gamma & \delta \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a & b-d & c & d \\ x & y-t & z & t \\ \alpha & \beta-\delta & \gamma & \delta \end{bmatrix},$$

where the second factor on the left-hand side is the transforming matrix $E_{42}(-1) \in M_4(K)$. However, if in some $(4 \times q)$-matrix it were necessary to replace the second row by its difference with the fourth row, then one would multiply on the left of A by $E_{24}(-1) \in M_4(K)$.

Transformation III. (*Row scaling*)

Multiplying the i-th row (i.e., each entry in that row) of a matrix A on the left by $\lambda \in K$ can be accomplished by multiplying the entire matrix A on the left by the matrix $E_{ii}(\lambda) \in M_p(K)$, which is obtained from the identity matrix I_p by replacing the i-th diagonal entry 1 with λ. For example:

$$E_{22}(3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}, E_{22}(3) \cdot \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 9 & 12 \\ 5 & 6 \end{bmatrix}.$$

Analogously, performing similar operations on the columns of A (instead of its rows) – where multiplication of entries by λ is now performed from the right – is equivalent to multiplying A on the right by the corresponding matrix $E_{jj}(\lambda) \in M_q(K)$, which is obtained from the identity matrix I_q by replacing the j-th diagonal entry 1 with λ. These column-transforming matrices differ from the row-transforming ones not only in size (when $p \neq q$) but also in that the roles of rows and columns are interchanged.

Remark 31. (*Elementary transformations*)

Transformations of types I-III applied to the rows or columns of a matrix are called *elementary transformations*.

- The corresponding elementary row transformation matrices are:

$$S\begin{bmatrix} 1 & 2 & \dots & p \\ i_1 & i_2 & \dots & i_p \end{bmatrix}, E_{ik}(\lambda) \in M_p(K), E_{ii}(\lambda) \in M_p(K),$$

where S represents a row permutation, $E_{ik}(\lambda)$ (with $i \neq k$) represents adding λ times row k to row i, and $E_{ii}(\lambda)$ represents multiplying row i by λ.

- The corresponding elementary column transformation matrices are:

$$S\begin{bmatrix} i_1 & i_2 & \dots & i_p \\ 1 & 2 & \dots & p \end{bmatrix}, E_{kj}(\lambda) \in M_q(K), E_{jj}(\lambda) \in M_q(K),$$

where S represents a column permutation, $E_{kj}(\lambda)$ (with $k \neq j$) represents adding λ times column j to column k, and $E_{jj}(\lambda)$ represents multiplying column j by λ.

These matrices are collectively called *elementary matrices*.

Let's illustrate, with an example, how a sequence of elementary transformations can be expressed as a product of their corresponding transformation matrices.

Example 34. Let the matrix

$$A = \begin{bmatrix} a & b & c & d \\ x & y & z & t \\ 1 & 2 & 3 & 4 \end{bmatrix}_{3\times 4} \quad \text{(over } \mathbb{Z}\text{)}$$

be subjected to the following elementary operations, in order:

1. Add the first row to the second row.
2. Swap the first and third columns.
3. Subtract twice the fourth column from the first column.
4. Multiply the second row by -1.

Only the first operation is performed directly on A; each subsequent operation is applied to the matrix obtained from the previous step.

(*Stepwise transformation*)

$$A \xrightarrow{\text{R2}\leftarrow\text{R2+R1}} \begin{bmatrix} a & b & c & d \\ x+a & y+b & z+c & t+d \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

$$\xrightarrow{\text{swap C1}\leftrightarrow\text{C3}} \begin{bmatrix} c & b & a & d \\ z+c & y+b & x+a & t+d \\ 3 & 2 & 1 & 4 \end{bmatrix}$$

$$\xrightarrow{\text{C1}\leftarrow\text{C1}-2\text{C4}} \begin{bmatrix} c-2d & b & a & d \\ z+c-2(t+d) & y+b & x+a & t+d \\ 3-2(4) & 2 & 1 & 4 \end{bmatrix}$$

$$\xrightarrow{\text{R2}\leftarrow-\text{R2}} \begin{bmatrix} c-2d & b & a & d \\ 2(t+d)-(z+c) & -(y+b) & -(x+a) & -(t+d) \\ -5 & 2 & 1 & 4 \end{bmatrix} = A'.$$

(*Matrix Factorization*)

Using elementary (row/column) transformation matrices and associativity of multiplication, the same process can be written as:

$$A \to E_{21}(1)\cdot A \to E_{21}(1)\,A\cdot S\begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{bmatrix} \to E_{21}(1)\,A\,S\begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{bmatrix}\cdot E_{41}(-2)$$

$$\to E_{22}(-1)\cdot E_{21}(1)\,A\,S\begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{bmatrix} E_{41}(-2)$$

$$= E_{22}(-1)E_{21}(1)\cdot A\cdot S\begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{bmatrix} E_{41}(-2) = UAV = A',$$

where

$$U = E_{22}(-1)E_{21}(1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$V = S\begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{bmatrix} E_{41}(-2) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -2 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -2 & 0 & 0 & 1 \end{bmatrix}.$$

(*Verification*)

$$UAV = UA\cdot V = \left(\begin{bmatrix} 1 & 0 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} a & b & c & d \\ x & y & z & t \\ 1 & 2 & 3 & 4 \end{bmatrix}\right)V$$

$$= \begin{bmatrix} a & b & c & d \\ -a-x & -b-y & -c-z & -d-t \\ 1 & 2 & 3 & 4 \end{bmatrix}\cdot\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -2 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} c + d(-2) & b & a & d \\ -(c+z) - (d+t)(-2) & -(b+y) & -(a+x) & -(d+t) \\ 3 + 4(-2) & 2\,1 & 4 & \end{bmatrix} = A'.$$

Transformation matrices U (for rows) and V (for columns) perform operations on matrix A that are more complex than elementary ones since they represent a sequence of several elementary operations. When constructing U, the order of the matrices in the product is reversed, while when forming B, the order is direct. It is important to note that during the process of altering a matrix directly, only the internal order of row operations and the internal order of column operations must be preserved. However, the two types of operations (rows and columns) may be interleaved in any way. This follows from the associativity of matrix multiplication (and operations in general – recall the multiplication of permutations). Thus, in Example 34, writing the transformation as $A \to UAV$ means that the row operations were applied first and the column operations second (with the order in each preserved). However, a different alternation of row and column operations had initially been specified. One can verify independently that if

$$U' = E_{12}(1) \cdot E_{22}(-1),\ V' = E_{41}(-2) \cdot S \begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{bmatrix},$$

then $U'AV \neq UAV \neq UAV'$.

It is clear that, in the general case, if the row operations correspond in order to the transforming matrices $U_1, U_2, \dots, U_m$, and the column operations correspond to $V_1, V_2, \dots, V_n$, then the original matrix A will finally transform into UAV, where $U = U_m U_{m-1} \dots U_1$ and $V = V_1 V_2 \dots V_n$.

Let's now consider rings of square matrices of order $p \in \mathbb{N}$ over an associative ring K with identity 1. Each ring $M_p(K)$ is itself associative and has an identity element $I = I_p$ (the zero element is the matrix $0 = (0_{ij})_p$). In what follows, for matrices in $M_p(K)$ with fixed p, we will omit the explicit mention of the dimension unless necessary. Our focus will be on the problem of invertibility of elements in the ring $M = M_p(K)$.

A matrix $A \in M$ is called *nonsingular* (or *nondegenerate*) if it is invertible; that is, if there exists a matrix A^{-1}, called the *inverse* of A, such that

$$A^{-1} \cdot A = A \cdot A^{-1} = I.$$

Conversely, a matrix is called *singular* (or *degenerate*) if it is noninvertible – that is, if not such inverse exists. Singular matrices include, in particular, the zero matrix.

According to the remark on the uniqueness of the inverse element in a monoid (Remark 3), if for a matrix $A = (a_{ij})$ there exist matrices X and Y such that $AX = YA = I$, then it follows that $X = Y = A^{-1}$; that is, A is nonsingular. Directly finding the inverse matrix $A^{-1} = (x_{ij})$ amounts to solving a system of p^2 linear equations in the p^2 unknowns $x_{11}, x_{12}, \dots, x_{pp}$. This is equivalent to a single matrix equation $(a_{ij}) \cdot (x_{ij}) = I$ or $(x_{ij}) \cdot (a_{ij}) = I$, with the same solution, even when the ring K is noncommutative. Thus, the nonsingularity of A is equivalent to the simultaneous solvability of the systems $(a_{ij}) \cdot (x_{ij}) = I$ and $(x_{ij}) \cdot (a_{ij}) = I$. If the existence of solutions in K for both systems is established, then in fact it suffices to solve only one of them. Conversely, if it is known that at least one of the systems is inconsistent, then A is a singular matrix.

Example 35. Over $K = \mathbb{Z}_6$ (the ring of residue classes modulo 6), consider the ring $M_2(\mathbb{Z}_6)$ of (2×2)-matrices. Although the elements 2 and 3 of $\mathbb{Z}_6$ are zero divisors (hence noninvertible), the matrix

$$\begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix} \in M_2(\mathbb{Z}_6)$$

is invertible: indeed

$$A^2 = \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix} \cdot \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 4+9 & 6+6 \\ 6+6 & 9+4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} (\text{mod } 6),$$

so $A^{-1} = A$.

By contrast, the matrix

$$B = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

is singular: if

$$BX = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I,$$

then comparing entries yields the system $2x_{11} = 1, 2x_{12} = 0, x_{21} = 0, x_{22} = 1$. The equation $2x_{11} = 1$ has no solution in $\mathbb{Z}_6$ because 2 is noninvertible, so no right inverse exists and B is singular.

As an exercise, investigate singularity/invertibility of the matrices

$$\begin{bmatrix} 2 & 3 \\ 2 & 3 \end{bmatrix}, \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}, \begin{bmatrix} 5 & 1 \\ 5 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 3 & 0 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 3 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 5 \\ 4 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 4 \\ 4 & 0 \end{bmatrix}$$

and find inverses for the nonsingular ones.

Example 36. Let A be a $(p \times p)$-matrix over a commutative ring K that is block/upper-triangular in the sense that all entries below the main diagonal are zero (so the matrix has diagonal entries $a_{11}, a_{22}, \dots, a_{pp}$). Fix k with $1 \leq k \leq p$:

$$A=\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1,k-1} & a_{1k} & a_{1,k+1} & \cdots & a_{1p} \\ 0 & a_{22} & a_{23} & \cdots & a_{2,k-1} & a_{2k} & a_{2,k+1} & \cdots & a_{2p} \\ 0 & 0 & a_{33} & \cdots & a_{3,k-1} & a_{3k} & a_{3,k+1} & \cdots & a_{3p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & a_{k-1,k-1} & a_{k-1,k} & a_{k-1,k+1} & \cdots & a_{k-1,p} \\ 0 & 0 & 0 & \cdots & 0 & a_{kk} & a_{k,k+1} & \cdots & a_{kp} \\ 0 & 0 & 0 & \cdots & 0 & 0 & a_{k+1,k+1} & \cdots & a_{k+1,p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & a_{p,k+1} & \cdots & a_{pp} \end{bmatrix}.$$

If the diagonal entry a_{kk} is *noninvertible in* K, then A is singular. Assume by contradiction that there exists $X=(x_{ij})$ with $XA=I$.

From the p^2 equations obtained by equating the entries of the matrix

$$XA=\begin{bmatrix} x_{11}a_{11} & x_{11}a_{12}+x_{12}a_{22} & \cdots & x_{11}a_{1k}+x_{12}a_{2k}+\cdots+x_{1k}a_{kk} \\ x_{21}a_{11} & x_{21}a_{12}+x_{22}a_{22} & \cdots & x_{21}a_{1k}+x_{22}a_{2k}+\cdots+x_{2k}a_{kk} \\ \cdots & \cdots & \cdots & \cdots \\ x_{k1}a_{11} & x_{k1}a_{12}+x_{k2}a_{22} & \cdots & x_{k1}a_{1k}+x_{k2}a_{2k}+\cdots+x_{kk}a_{kk} \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

componentwise to those of the matrix I, we obtain

$x_{11}a_{11}=1$,

$x_{21}a_{11}=0$, $x_{21}a_{12}+x_{22}a_{22}=1$,

$x_{31}a_{11}=0$, $x_{31}a_{12}+x_{32}a_{22}=0$, $x_{31}a_{13}+x_{32}a_{23}+x_{33}a_{33}=1$,

……

$x_{k1}a_{11}=0$, $x_{k1}a_{12}+x_{k2}a_{22}=0$, $x_{k1}a_{1,k-1}+\cdots+x_{k,k-1}a_{k-1,k-1}=0$,

$$x_{k1}a_{1k}+\cdots+x_{k,k-1}a_{k-1,k}+x_{kk}a_{kk}=1.$$

The first equation implies a_{11} is invertible so the remaining equalities in the first column force $x_{21}=x_{31}=\cdots=x_{k1}=0$.

Now compare entries in the second column using the already-known vanishing of the x_{i1}:

$$x_{21}a_{12}+x_{22}a_{22}=1,\ x_{32}a_{22}=0,\ \dots,\ x_{k2}a_{22}=0,$$

and we deduce $x_{22}a_{22}=1$. Hence a_{22} is invertible and $x_{32}=\cdots=x_{k2}=0$. Continuing this way, we find successively that $a_{33},\dots,a_{k-1,k-1}$ are invertible and many x_{ij} vanish, until finally the equation coming from the (k,k)-entry reads

$$x_{kk}a_{kk}=1,$$

which contradicts the assumed noninvertibility of a_{kk}. Thus, no such X exists and A is singular.

Further important examples of inversion are the transformation matrices already known to us.

Example 37. For any permutation $i_1,i_2,\dots,i_p$ of the numbers $1,2,\dots,p$, the matrix

$$S\begin{bmatrix} 1 & 2 & \dots & p \\ i_1 & i_2 & \dots & i_p \end{bmatrix}$$

is nonsingular. Its inverse can be verified directly and is given by

$$S^{-1}\begin{bmatrix} i_1 & i_2 & \dots & i_p \\ 1 & 2 & \dots & p \end{bmatrix},$$

where the unit of the 1st column lies in the first row, the unit of the 2nd column lies in the second row, and so forth. Indeed, it is clear that to restore the original matrix A from the product SA, one must move the i_1-th row back to the first position, the i_2-th row to the second, and so on. Let's make Example 37 concrete with $p=3$.

Take the numbers 1,2,3. Suppose we choose the permutation $(i_1,i_2,i_3)=(2,3,1)$.

(*Construct S*)

$$S\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}.$$

This matrix encodes:

- the 1st row of A goes to position 2,
- the 2nd row goes to position 3,
- the 3rd row goes to position 1.

So, multiplying SA permutes the rows of A.

(*Compute the inverse* S^{-1})

$$S^{-1}\begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \end{bmatrix}.$$

This says:

- move row 2 back to row 1,
- move row 3 back to row 2,
- move row 1 back to row 3.

(*Test with a sample matrix*) Let

$$A=\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}.$$

Then

$$SA = \begin{bmatrix} e & f \\ a & b \\ c & d \end{bmatrix}.$$

Here the rows are permuted: row 3 is now 1st, row 1 is 2nd, row 2 is 3rd. Now, apply S^{-1}:

$$S^{-1}(SA) = A.$$

So, the permutation is perfectly undone.

(*Key points used*)

- S permutes the rows of matrix A according with the chosen permutation.
- S^{-1} restores the original order.

Example 38. For any $\lambda \in K$, the inverse matrix of $E_{ik}(\lambda)$ is $E_{ik}(-\lambda)$. For instance, when $p = 4, i = 2$, and $k = 3$, we have

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \lambda & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -\lambda & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -\lambda & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \lambda & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This is natural: if the matrix A' is obtained from A by adding the k-th row, multiplied by λ, to the i-th row, then to recover A from A', one must add the k-th row, multiplied by $-\lambda$, to the i-th row.

Example 39. The matrix $E_{ii}(\lambda)$ is nonsingular if and only if λ is an invertible element of the ring K (proof left to the reader). In that case, the inverse is given by $E_{ii}(\lambda^{-1})$.

Analogous statements hold for matrices which, when multiplied on the right, transform the columns of a given matrix A.

Thus, in principle, the problem of inverting matrices $A \in M_p(K)$ is solved. However, the explicit computation of A^{-1} (or the determination of whether A is singular) is, in general, rather labor-intensive. The process can be simplified by assuming that every nonzero element of K is invertible; that is, by assuming K is a skew field. To proceed, we require the matrices of elementary transformations together with Example 36 (which serves as a lemma). In that lemma, the condition that an element a_{kk} is noninvertible is replaced with the simpler condition $a_{kk} = 0$. In practice, K is usually even a field, but commutativity of multiplication is never required.

Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix}, A \in M_p(K),$$

where K is a skew field, $p \in \mathbb{N}$. We describe a process which, if it terminates at some stage, shows immediately that the matrix A is singular, and which, if carried out to the end, either yields the inverse A^{-1}, or shows that it does not exist.

(*First stage*) If $a_{11} = a_{21} = \cdots = a_{p1} = 0$, the *process terminates at the first stage*, and A is singular. Otherwise, choose i such that $a_{i1} \neq 0$. Multiply the i-th row on the left by a_{i1}^{-1} and move it to the first row (the other rows may be rearranged for convenience). The corresponding transformation matrices are $E_{ii}(a_{i1}^{-1})$ and $S\begin{bmatrix} i & \cdots \\ 1 & \ldots \end{bmatrix}$, so that

$$A' = \begin{bmatrix} 1 & a'_{12} & \cdots & a'_{1p} \\ a'_{21} & a'_{22} & \cdots & a'_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a'_{p1} & a'_{p2} & \cdots & a'_{pp} \end{bmatrix} = S\begin{bmatrix} i & \cdots \\ 1 & \ldots \end{bmatrix} E_{ii}(a_{i1}^{-1})\, A.$$

If $a'_{21} = \cdots = a'_{p1} = 0$, the first stage is complete. Otherwise, for each $r \geq 2$ with $a'_{r1} \neq 0$, add to the r-th row the first row multiplied on the left by $-a'_{r1}$. The corresponding matrices are $E_{r1}(-a'_{r1})$. At the end of the first stage, the matrix has the form

$$A'' = \begin{bmatrix} 1 & a''_{12} & \cdots & a''_{1p} \\ 0 & a''_{22} & \cdots & a''_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a''_{p2} & \cdots & a''_{pp} \end{bmatrix}.$$

(*Second stage*) If $a''_{22} = \cdots = a''_{p2} = 0$, the *process terminates at the second stage*, and A is singular. Otherwise, as in the first stage, a leading 1 is produced in position $(2, 2)$, the entries are cleared below it, and finally a''_{12} is eliminated (if nonzero) by adding to the first row the second row multiplied on the left by $-a''_{12}$ (This step is skipped if $a''_{12} = 0$).

At the end of the second stage (if the process has not terminated earlier), the matrix has the form

$$A''' = \begin{bmatrix} 1 & 0 & a'''_{13} & \cdots & a'''_{1p} \\ 0 & 1 & a'''_{23} & \cdots & a'''_{2p} \\ \vdots & \vdots & \vdots\ddots & & \vdots \\ 0 & 0 & a'''_{p3} & \cdots & a'''_{pp} \end{bmatrix}$$

(*General stage*) Proceeding similarly, at each stage k either:

1. we obtain a matrix $\tilde{A}$ whose all entries in column k below the pivot position vanish, in which case the process terminates and A is singular; or

2. a pivot element is produced in position (k, k), zeros are cleared below it, and the column entries above it are eliminated.

(*Termination*) Two cases are possible:

1. If all p stages of the process are completed, the result is the identity matrix, and the product of the elementary matrices used gives A^{-1}.

2. If the process terminates earlier, the resulting matrix has a zero column in the reduced block, implying that A is singular.

Remarks 32. We have described a row-reduction procedure (*Gaussian elimination* with left multiplications) that either produces a factorization from which A^{-1} can be computed, or else terminates earlier, proving that A is singular.

Notation: all scalar inverses and scalar multiplications of rows are performed on the left (since K may be noncommutative). Elementary row operations are represented by left multiplication with elementary matrices.

First case. $I = U_m U_{m-1} \dots U_1 A$, where $U_1, U_2, \dots, U_m$ are the matrices corresponding to the elementary row transformations performed at all p stages, in order: first the operations of the initial stage (in sequence), then those of the second stage, and so on. Thus, $U = U_m U_{m-1} \dots U_1$ is the only candidate for the role of A^{-1} since it satisfies the relation $UA = I$. However, it is still necessary to verify directly that $AU = I$. This verification is essential, because $UA = I$ only guarantees that U is a *left inverse* of A, but not necessarily the true inverse. Avoiding this check is inadvisable: even if (under certain additional assumptions on K and A) the verification were theoretically unnecessary, in practice it remains crucial since the likelihood of an error in at least one transformation is fairly high.

Second case. If the process terminates at the k-th stage ($1 \le k \le p$), the matrix $\tilde{A}$ is singular by Example 36. We now show that the original matrix A must also be singular.

Indeed, $\tilde{A} = U_m U_{m-1} \dots U_1 A$, where $U_1, U_2, \dots, U_m$ are the matrices corresponding to the elementary transformations that convert A into $\tilde{A}$. Each U_i is nonsingular. If A were nonsingular, then by the theorem on the invertibility of a product in a monoid, the matrix $\tilde{A}$ would also be nonsingular – a contradiction.

Example 40. (*Matrix Inversion Example*) The procedure for inverting the matrix A, although already described and justified, has not yet been presented in a concise form. Before doing so, let's consider an illustrative example.

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 3 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}, (K = \mathbb{Q}, K = \mathbb{R} \text{ or } K = \mathbb{C}; p = 3).$$

We perform a sequence of elementary row transformations. Above each arrow, we place the symbol of the corresponding transformation matrix. The reduction of A to the identity matrix is carried out as follows:

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 3 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix} \xrightarrow{E_{33}(1/2)} \begin{bmatrix} 0 & 1 & 1 \\ 3 & 0 & 1 \\ 1 & 1/2 & 0 \end{bmatrix} \xrightarrow{S\begin{bmatrix} 3 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix}} \begin{bmatrix} 1 & 1/2 & 0 \\ 0 & 1 & 1 \\ 3 & 0 & 1 \end{bmatrix} \xrightarrow{E_{31}(-3)} \begin{bmatrix} 1 & 1/2 & 0 \\ 0 & 1 & 1 \\ 0 & -3/2 & 1 \end{bmatrix}$$

$$\xrightarrow{E_{32}(3/2)} \begin{bmatrix} 1 & 1/2 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 5/2 \end{bmatrix} \xrightarrow{E_{12}(-1/2)} \begin{bmatrix} 1 & 0 & -1/2 \\ 0 & 1 & 1 \\ 0 & 0 & 5/2 \end{bmatrix} \xrightarrow{E_{33}(2/5)} \begin{bmatrix} 1 & 0 & -1/2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{E_{23}(-1)} \begin{bmatrix} 1 & 0 & -1/2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow{E_{13}(1/2)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I.$$

Since each elementary transformation is simultaneously applied to the identity matrix I, the inverse A^{-1} is obtained as the product

$$U = E_{13}(1/2) E_{23}(-1) E_{33}(2/5) E_{12}(-1/2) E_{32}(3/2) E_{31}(-3) S\begin{bmatrix} 3 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix} E_{32}(1/2).$$

Carrying out the matrix multiplications step by step:

$$U = \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2/5 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1/2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3/2 & 1 \end{bmatrix} \cdot$$

$$\cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/2 \end{bmatrix} = \begin{bmatrix} -1/5 & 1/5 & 1/5 \\ 2/5 & -2/5 & 3/5 \\ 3/5 & 2/5 & -3/5 \end{bmatrix}.$$

As an exercise, one may explicitly verify the product of all factors (instead of computing them sequentially, as above), which replaces the need for seven smaller examples.

(*Verification*)

$$AU = \begin{bmatrix} 0 & 1 & 1 \\ 3 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} -1/5 & 1/5 & 1/5 \\ 2/5 & -2/5 & 3/5 \\ 3/5 & 2/5 & -3/5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Thus, $U = A^{-1}$ since both $AU = UA = I$.

Example 41. (*Matrix inversion using the method of elementary transformations*) Having understood the procedure described earlier, it is now easy to follow an elegant implementation (due to D.N. Krylov), which avoids the need to explicitly write out the transformation matrices.

Given a matrix $A = (a_{ij}) \in M_p(K)$, we append the identity matrix $I = I_p$ to its right, thereby forming a $(p \times 2p)$-matrix:

$$\left[\begin{array}{cccc|cccc} a_{11} & a_{12} & \cdots & a_{1p} & 1 & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2p} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} & 0 & 0 & \cdots & 1 \end{array}\right]. \qquad (13)$$

Applying a sequence of elementary row transformations to this augmented matrix, we transform its left half into the identity matrix. Simultaneously, the right hand becomes a matrix U satisfying $UA = I$ (why?). To complete the argument, one verifies directly that $AU = I$, and hence $U = A^{-1}$.

This technique is called *inversion by the method of elementary transformations* and is conventionally denoted by

$$[A|I] \xrightarrow{E_1, E_2, \ldots, E_k} [I|A^{-1}].$$

(*Illustrative Computation*) Consider the matrix A of Example 40.

$$\left[\begin{array}{ccc|ccc} 0 & 1 & 1 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 & 1 & 0 \\ 2 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 0 & 1 & 1 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1/2 & 0 & 0 & 0 & 1/2 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 & 1 & 0 \end{array}\right] \to$$

$$\to \left[\begin{array}{ccc|ccc} 1 & 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & -3/2 & 1 & 0 & 1 & -3/2 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 5/2 & 3/2 & 1 & -3/2 \end{array}\right] \to$$

$$\to \left[\begin{array}{ccc|ccc} 1 & 0 & -1/2 & -1/2 & 0 & 1/2 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 5/2 & 3/2 & 1 & -3/2 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & -1/2 & -1/2 & 0 & 1/2 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 3/5 & 2/5 & -3/5 \end{array}\right] \to$$

$$\to \left[\begin{array}{ccc|ccc} 1 & 0 & -1/2 & -1/2 & 0 & 1/2 \\ 0 & 1 & 0 & 2/5 & -2/5 & 3/5 \\ 0 & 0 & 1 & 3/5 & 2/5 & -3/5 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -1/5 & 1/5 & 1/5 \\ 0 & 1 & 0 & 2/5 & -2/5 & 3/5 \\ 0 & 0 & 1 & 3/5 & 2/5 & -3/5 \end{array}\right].$$

Thus, the right-hand side is precisely A^{-1}, and the relation $AU = I$ has already been verified.

(*Another approach*) Instead of working with rows, one can form a $(2p \times p)$-matrix

$$\begin{bmatrix} A \\ I_p \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

and now apply elementary column transformations instead of row transformations. The goal is to transform the top block into I_p; simultaneously, the bottom block will become a matrix V such that $AV = I$. It then remains to check directly that also $VA = I$. We suggest carrying this out independently for the same (3×3)-matrix A from the previous example, and verifying that the resulting matrix V coincides with the earlier inverse U. Elementary transformations with columns can also be performed on a matrix written in augmented form as in (13), provided that every operation on the columns of the left block is simultaneously applied to the corresponding columns of the right block.

Question. The idea that arises is to transform both the rows and columns of matrix (13), with the goal of reducing the left half to the identity matrix. For example,

$$\left[\begin{array}{ccc|ccc} 0 & 1 & 1 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 & 1 & 0 \\ 2 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \xrightarrow{\text{swap columns: } S\begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}} \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 3 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 1 & 0 \end{array}\right]$$

$$\xrightarrow{C_2 \leftarrow C_2 - C_1} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 3 & 1 & -1 & 0 \\ 1 & -1 & 2 & 0 & 1 & 0 \end{array}\right] \xrightarrow{R_3 \leftarrow R_3 - R_1} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 3 & 1 & -1 & 0 \\ 0 & -1 & 2 & 0 & 1 & -1 \end{array}\right]$$

$$\xrightarrow{R_3 \leftarrow R_3 + R_2} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 3 & 1 & -1 & 0 \\ 0 & 0 & 5 & 1 & 0 & -1 \end{array}\right] \xrightarrow{R_3 \leftarrow \frac{1}{5}R_3} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 3 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1/5 & 0 & -1/5 \end{array}\right]$$

$$\xrightarrow{R_2 \leftarrow R_2 - 3R_3} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2/5 & -1 & 3/5 \\ 0 & 0 & 1 & 1/5 & 0 & -1/5 \end{array}\right].$$

Thus, the candidate for the inverse is

$$W = \begin{bmatrix} 0 & 0 & 1 \\ 2/5 & -1 & 3/5 \\ 1/5 & 0 & -1/5 \end{bmatrix}.$$

However, direct verification shows that $W \neq A^{-1}$. *What went wrong with this approach*?

One of the main applications of matrix inversion is the solution of systems of linear equations

$$\left.\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p = b_2 \\ \cdots\cdots\cdots\cdots \\ a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pp}x_p = b_2 \end{array}\right\},$$

where the coefficients a_{ij}, the unknowns x_j, and the constant terms b_j belong to a certain field P. Defining

$$A = \left(a_{ij}\right)_{p\times p} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix}, X = \left(x_j\right)_{p\times 1} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}, B = \left(b_j\right)_{p\times 1} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix},$$

we can rewrite the system compactly as

$$AX = B. \quad (14)$$

A *solution* is not only the process of determining the values of the unknowns $x_1, x_2, \ldots, x_p$, but also the ordered tuple of values $\check{x}_1, \check{x}_2, \ldots, \check{x}_p \in P$. Substituting these values into the system gives true equalities in P. Equivalently, a solution (or *root*) of (14) is a vector $\check{X} = \left(\check{x}_j\right)_{p\times 1}$ for which $A\check{X} = B$.

Claim 3. *If A is a nonsingular matrix, then equation* (14) *has a unique solution $\check{X} = A^{-1}B$.* •

Indeed, assuming first that some solution $\check{X}$ exists, we obtain $\check{X} = A^{-1}B$ from the equality $A\check{X} = B$ by multiplying by A^{-1} on the left, and then by substituting into (14) we verify that this unique candidate is indeed a solution (compare with Theorem 4 on the unique solvability of equations in a group).

Corollary 8. *If $b_1 = b_2 = \cdots = b_p = 0$* (in this case, the system of equations is called *homogeneous*) *and the matrix A is nonsingular, then the only solution to the system is the trivial* (*zero*) *solution $\check{x}_1 = \check{x}_2 = \cdots = \check{x}_p = 0$.* •

Remark 33. As follows from the above, for a homogeneous system

$$\left.\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p = 0 \\ \cdots\cdots\cdots\cdots \\ a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pp}x_p = 0 \end{array}\right\}$$

to admit nontrivial solutions (that is, when at least one $\check{x}_j \neq 0$), it is necessary that the coefficient matrix $A = \left(a_{ij}\right)_{p\times p}$ be singular. In fact, this condition is also sufficient: a homogeneous system of p equations in p unknowns with a singular matrix necessarily possesses nontrivial solutions. The proof, however, is not entirely straightforward and will be given later in the general theory of linear systems (semester 2).

Exercise. Apply the matrix approach to the systems of equations from Exercises $4 - 4''$ of the introductory lecture (see Part I of this course), and also to the corresponding homogeneous systems obtained by replacing all constants on the right-hand sides with zeros.

MATERIAL FOR INDEPENDENT STUDY.

(*based on G.E. Shilov, Mathematical Analysis, Part I: Finite-Dimensional Linear Spaces, Moscow, Nauka*, 1969)

- Matrix multiplication (over an arbitrary ring), including block partitioning.
- Matrix transposition; the transposition theorem for the product of matrices over a commutative ring.

These topics are neither difficult nor lengthy. Their distribution near the end of the examination, and their inclusion in the exam syllabus, will serve as a useful test of whether students have acquired at least minimal skills in the independent study of mathematical literature during the semester.

CONCLUSION

Only in this part of the course – intended for the entire first semester – does our teaching methodology differ significantly from the traditional one. Beginning in the second semester, one may, for example, adopt the following textbook as a basis:

N. V. Efimov, E. R. Rozendorn, *Linear Algebra and Multidimensional Geometry*, *Moscow*, *Nauka*, 1970,

supplemented with the theory of determinants from either G. E. Shilov or A. G. Kurosh.

This consideration motivated us to publish the initial part of this algebra course separately, without waiting for the completion of the subsequent volumes.

CONCLUDING NOTE

(Carlos E. Frasser)

Professor Zykov's approach to algebra as a scientific discipline emphasizes the independent development of fundamental concepts and the gradual transition to more general results. I would add that this methodology remains pedagogically valuable today: it enables students not only to master computational techniques but also to develop the capacity to reason abstractly about groups, rings, fields, polynomials, and matrix theory.

In this spirit, the exercises and independent study materials serve as a crucial bridge between formal theory and personal discovery. My experience confirms that such a balance between guidance and autonomy is essential for fostering genuine mathematical maturity in students.

References

1. Frasser C.E. *A Course of Algebra* (*Part I*). 2024. ArXiv:2411.11873.