Chapter 2

Dramatis Personae

1 Vector Spaces

In four words, the realm of linear algebra can be described as *geometry* of vector spaces.

The Axioms

By definition, a **vector space** is a set, equipped with operations of **addition** and **multiplication by scalars** which are required to satisfy certain **axioms**.

The set will be denoted here by $\mathcal{V}$, and its elements referred to as **vectors**. Here are the axioms.

(i) The sum of two vectors $\mathbf{u}$ and $\mathbf{v}$ is a vector (denoted $\mathbf{u}+\mathbf{v}$); the result of multiplication of a vector $\mathbf{v}$ by a scalar λ is a vector (denoted $\lambda\mathbf{v}$).

(ii) The addition of vectors is commutative and associative:

$$\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u},\ (\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})\ \text{ for all } \mathbf{u},\mathbf{v},\mathbf{w}\in\mathcal{V}.$$

(iii) There exists the **zero vector** (denoted by $\mathbf{0}$) such that

$$\mathbf{v}+\mathbf{0}=\mathbf{0}+\mathbf{v}=\mathbf{v}\ \text{ for every } \mathbf{v}\in\mathcal{V}.$$

(iv) For every vector $\mathbf{u}$ there exists the **opposite** vector, denoted by $-\mathbf{u}$, such that

$$-\mathbf{u}+\mathbf{u}=\mathbf{0}.$$

(v) Multiplication by scalars is distributive: For all vectors $\mathbf{u}, \mathbf{v}$ and scalars λ, μ we have

$$(\lambda + \mu)(\mathbf{u} + \mathbf{v}) = \lambda\mathbf{u} + \lambda\mathbf{v} + \mu\mathbf{u} + \mu\mathbf{v}.$$

(vi) Multiplication by scalars is associative in the following sense: For every vector $\mathbf{u}$ and all scalars λ, μ we have:

$$(\lambda\mu)\mathbf{u} = \lambda(\mu\mathbf{u}).$$

(vii) Multiplication by scalars 0 and 1 acts on every vector $\mathbf{u}$ as

$$0\mathbf{u} = \mathbf{0}, \quad 1\mathbf{u} = \mathbf{u}.$$

We have to add to this definition the following comment about **scalars**. Taking one of the sets $\mathbb{R}$ or $\mathbb{C}$ of real or complex numbers on the role of scalars one obtains the definition of **real** vector spaces or **complex** vector spaces. In fact these two choices will suffice for all our major goals. The reader may assume that we use the symbol $\mathbb{K}$ to cover both cases $\mathbb{K} = \mathbb{R}$ and $\mathbb{K} = \mathbb{C}$ in one go.

On the other hand, any **field** $\mathbb{K}$ qualifies on the role of the set of scalars, and this way one arrives at the notion of $\mathbb{K}$-vector spaces. Our general theory works *verbatim* in this more abstract setting. Therefore, after the following (optiona) digression on fields, the reader may choose to assume that $\mathbb{K}$ stands for an arbitrary field of scalars.

Fields

By a **field** one means a set $\mathbb{K}$ equipped with two operations: addition and multiplication. Both are assumed to be commutative and associative, and satisfying the distributive law: $a(b + c) = ab + ac$. Besides, it is required that there exist elements 0 and $1 \neq 0$ such that $a + 0 = a$ and $1a = a$ for all $a \in \mathbb{K}$. Then, it is required that every $a \in \mathbb{K}$ has the opposite $-a$ such that $-a + a = 0$, and that every *non-zero* $a \in \mathbb{K}$ has its *inverse* a^{-1} such that $a^{-1}a = 1$. To the examples of fields $\mathbb{C}$ and $\mathbb{R}$, we can add (omitting many other available examples): the field $\mathbb{Q}$ of rational numbers; the field $\mathcal{A} \subset \mathbb{C}$ of all **algebraic numbers** (i.e. roots of polynomials in one variable with rational coefficients); the field $\mathbb{Z}_p$ of integers modulo a given prime number p (see Exercises). For instance, the set $\mathbb{Z}_2 = \{0, 1\}$ of remainders modulo 2 with the usual arithmetic of remainders ($0+0 = 0 = 1+1$, $0+1 = 1 = 1+0$, $0 \cdot 0 = 1 \cdot 0 = 0 \cdot 1 = 0$, $1 \cdot 1 = 1$) can be taken on the role of scalars. This gives rise to the definition of $\mathbb{Z}_2$-vector spaces useful in computer science and logic.

To reiterate: it is essential that division by all non-zero scalars is defined. Therefore the set $\mathbb{Z}$ of all integers and the set $\mathbb{F}[x]$ of all polynomials in one indeterminate x with coefficients in a field $\mathbb{F}$ are not fields, and do not qualify on the role of scalars in the definition of vector spaces, because the division is not always possible. However the field $\mathbb{Q}$ of all rational numbers and the field $\mathbb{F}(x)$ of all rational functions with coefficients in a field $\mathbb{F}$ are O.K.

EXERCISES

103. Can a field have only one element? ϟ

104. Verify that $\mathbb{Z}_2$ is a field.

105. Prove uniqueness of the zero element, the unit element, the opposite and the inverse to a given element. ϟ

106. Given two integers a and $b > 0$, let q and r be respectively the quotient and remainder from the division of a by b, i.e. $a = qb + r$, where $0 \leq r < b$. Show that the set of common divisors of a and b coincides with the set of common divisors of b and r.

107. Prove that for any two integers a and b, there exists a non-negative integer d (often denoted (a, b) and called the **greatest common divisor** of a and b) such that the set of common divisors of a and b coincides with the set of all divisors of d. ✓

108. Prove that the greatest common divisor (a, b) of two integers a and b is representable in the form $d = ka + lb$, where k, l are some integers. ✓

109. Let n be a positive integer. Call integers a and b **congruent modulo** n (and write $a \cong b \mod n$) if $a - b$ is divisible by n. Prove that if $a \cong b \mod n$ and $a' \cong b' \mod n$ then $a+a' \cong b+b' \mod n$ and $ab \cong a'b' \mod n$.

110. Denote by $\mathbb{Z}_n$ the set of congruence classes of integers modulo n. Show that addition and multiplication of integers descends to the addition and multiplication on $\mathbb{Z}_n$. ϟ

111. How many elements does $\mathbb{Z}_n$ have? ✓

112. Find all invertible elements and their multiplicative inverses in $\mathbb{Z}_5$.

113. The same for $\mathbb{Z}_8$. ✓

114. Prove that the congruence class of an integer a has a multiplicative inverse in $\mathbb{Z}_n$ if and only if a is relatively prime to n. ϟ

115. Prove that $\mathbb{Z}_n$ is a field if and only if n is prime.

116. By definition, a **homomorphism** $f : \mathbb{F} \to \mathbb{K}$ is a map respecting the operations: $f(a + b) = f(a) + f(b)$ and $f(ab) = f(ab) = f(a)f(b)$ for all $a, b \in \mathbb{F}$. Prove that a non-zero homomorphism between two fields is necessarily injective. ϟ

117. Show that numbers of the form $a + b\sqrt{2}$, where $a, b \in \mathbb{Q}$, form a subfield in $\mathbb{R}$.

118. Prove that every field $\mathbb{K}$ contains a unique subfield isomorphic to $\mathbb{Q}$ or one of $\mathbb{Z}_p$, where p is prime. (One calls $\mathbb{K}$ a fields of **characteristic** p in the latter case, and of characteristic 0 in the former.) ϟ

119.* Show that there is a field $\mathbb{F}_4$ of four elements. ϟ

120. Find all roots of the polynomial $x^2 + x + 1$ in $\mathbb{F}_4$. ϟ

A Proof

From the axioms, all further properties of vectors are derived. As an illustration, let us prove that

$$-\mathbf{u} = (-1)\mathbf{u} \quad \text{for all } \mathbf{u}.$$

Indeed, $1 + (-1) = 0$. By the axiom (vii), $0\mathbf{u} = \mathbf{0}$, and hence $\mathbf{0} = (1 + (-1))\mathbf{u} = \mathbf{u} + (-1)\mathbf{u}$, where we applied the distributive law (v) and $1\mathbf{u} = \mathbf{u}$ from (vii). Now we add $-\mathbf{u}$ to both sides, to find $-\mathbf{u} + \mathbf{0} = -\mathbf{u}$ (by (iii)) on the left, and

$$-\mathbf{u} + (\mathbf{u} + (-1)\mathbf{u}) = (-\mathbf{u} + \mathbf{u}) + (-1)\mathbf{u} = \mathbf{0} + (-1)\mathbf{u} = (-1)\mathbf{u}$$

(consecutively by (ii), (iv), and (iii)) on the right. Thus $-\mathbf{u} = (-1)\mathbf{u}$.

Such derivations might be a fun and useful game, but this is not the purpose of introducing the notion of a vector space, nor are the axioms "absolute truth accepted without proof." The abstract axiomatic definition of vector spaces is in fact a unification device: it allows one to study many examples at once without the need to reiterate the same derivations. Yet, to apply general conclusions, the validity of the axioms needs to be verified in each example separately. The axioms are modeleled, of course, on the properties of geometric vectors. However, to justify the axiomatic approach, at least one more example is needed. It will be provided shortly.

EXERCISES

121. Prove that in a vector space, the zero vector is unique. ϟ

122. Prove that the opposite of each vector is unique. ϟ

123. Give an example of a "vector space" that satisfies all axioms except the last one: $1\mathbf{u} = \mathbf{u}$ for all $\mathbf{u}$. ϟ

124. Prove that the axiom: $0\mathbf{u} = \mathbf{0}$ for all $\mathbf{u}$, in the definition of vector spaces is redundant, i.e. can be derived from the remaining axioms. ϟ

$\mathbb{K}^n$

Let n be a non-negative integer. The set of all *columns* of height n whose entries are elements of $\mathbb{K}$ is denoted by $\mathbb{K}^n$ (i.e. $\mathbb{R}^n$ is the set of such columns of real numebrs, and $\mathbb{C}^n$ of complex). It is equipped with the operations of *termwise* addition and multiplication by scalars:

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} := \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}, \quad \lambda \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} := \begin{bmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{bmatrix}.$$

We leave it as an exercise for the reader to verify that the set $\mathbb{K}^n$ equipped with these operation satisfies all the axioms of a $\mathbb{K}$-vector space. It is called the **standard coordinate space**, (or **n-space**, its elements $\mathbf{x}, \mathbf{y} \in \mathbb{K}^n$ are called **coordinate vectors** (or **n-vectors**) , and their components $x_i, y_i \in \mathbb{K}$ **coordinates** of the vectors. In paticular, the vector all of whose coordinates are zeros plays the role of the vector $\mathbf{0}$.

Subspaces

A *non-empty* subset $\mathcal{W} \subset \mathcal{V}$ in a vector space $\mathcal{V}$ is called a **linear subspace** (or **vector subspace**, or simply **subspace**), if it is closed with respect to the operations of addition of vectors and multiplication by scalars, i.e. for all $\mathbf{u}, \mathbf{v} \in \mathcal{W}$ and all $\lambda, \mu \in \mathbb{K}$, we have $\lambda\mathbf{u} + \mu\mathbf{v} \in \mathcal{W}$. In this case, $\mathcal{W}$ forms a vector space on its own. Indeed, using the assumption that $\mathcal{W}$ is non-empty, we can pick an element $\mathbf{w}$ in it, multiply it by scalar 0, and conclude that vector $\mathbf{0} = 0\mathbf{w}$ lies in $\mathcal{W}$. Furthermore, for any $\mathbf{w} \in \mathcal{W}$, we find that $-\mathbf{w} = (-1)\mathbf{w}$ also lies in $\mathcal{W}$. Consequently, $\mathcal{W}$ forms a vector space on its own. All other axioms are satisfied for elements of $\mathcal{W}$ bacause they are satisfied for all elements in the ambient space $\mathcal{V}$.

In practice, the spaces $\mathbb{K}^n$ and their linear subspaces form a supply of examples of vector spaces suficient for our goals.

EXERCISES

125. Show that in $\mathbb{K}[x]$, polynomials of degree n do not form a subspace, but polynomials of degree $\leq n$ do.

126. Prove that intersection of subspaces is a subspace.

127. Show that the union of two subspaces is a subspace if and only if one of the subspaces contains the other.

Morphisms

The modern math ideology requires that mathematical entities are organized into *categories*. This means that in addition to specifying *objects* of interest, one should also specify *morphisms*, i.e. maps between them. The **category of vector spaces** is obtained by taking vector spaces for objects, and linear maps for morphisms.

By definition, a function $A : \mathcal{V} \to \mathcal{W}$ from a vector space $\mathcal{V}$ to a vector space $\mathcal{W}$ is called a **linear map** if it *respects* the operations with vectors, i.e. if it maps linear combinations of vectors to linear combinations of their images with the same coefficients:

$$A(\lambda\mathbf{u} + \mu\mathbf{v}) = \lambda A\mathbf{u} + \mu A\mathbf{v} \quad \text{for all } \mathbf{u}, \mathbf{v} \in \mathcal{V} \text{ and } \lambda, \mu \in \mathbb{K}.$$

With a linear map $A : \mathcal{V} \to \mathcal{W}$, one associates two subspaces, one in $\mathcal{V}$, called the **null space**, or the **kernel** of A and denoted $\operatorname{Ker} A$, and the other in $\mathcal{W}$, called the **range** of A and denoted $A(\mathcal{V})$:

$$\operatorname{Ker} A := \{\mathbf{v} \in \mathcal{V} : A\mathbf{v} = \mathbf{0}\},$$
$$A(\mathcal{V}) := \{\mathbf{w} \in \mathcal{W} : \mathbf{w} = A\mathbf{v} \text{ for some } \mathbf{v} \in \mathcal{V}\}.$$

A linear map is **injective**, i.e. maps different vectors to different ones, exactly when its kernel is trivial. Indeed, if $\operatorname{Ker} A \neq \{\mathbf{0}\}$, then it contains non-zero vectors mapped to the same point $\mathbf{0}$ in $\mathcal{W}$ as $\mathbf{0}$ from $\mathcal{V}$. This makes the map A non-injective. *Vice versa*, if A is non-injective, i.e. if $A\mathbf{v} = A\mathbf{v}'$ for some $\mathbf{v} \neq \mathbf{v}'$, then $\mathbf{u} = \mathbf{v} - \mathbf{v}' \neq \mathbf{0}$ lies in $\operatorname{Ker} A$. This makes the kernel nontrivial.

When the range of a map is the whole target space, $A(\mathcal{V}) = \mathcal{W}$, the map is called **surjective**. If a linear map $A : \mathcal{V} \to \mathcal{W}$ is **bijective**, i.e. both injective and surjective, it establishes a one-to-one correspondence between $\mathcal{V}$ and $\mathcal{W}$ in a way that respects vector operations. Then one says that A establishes an **isomorphism** between the vector spaces $\mathcal{V}$ and $\mathcal{W}$. Two vector spaces are called **isomorphic** (written $\mathcal{V} \cong \mathcal{W}$) if there exists an isomorphism between them.

Direct Sums

The are many constructions of making new vector spaces from given ones. Here is one which will be used most frequently (see Figure 23).

Given two vector spaces $\mathcal{V}$ and $\mathcal{W}$, their **direct sum** $\mathcal{V} \oplus \mathcal{W}$ is defined as the set of all ordered pairs $(\mathbf{v}, \mathbf{w})$, where $\mathbf{v} \in \mathcal{V}$, $\mathbf{w} \in \mathcal{W}$, equipped with the component-wise operations:

$$\lambda(\mathbf{v}, \mathbf{w}) = (\lambda\mathbf{v}, \lambda\mathbf{w}), \quad (\mathbf{v}, \mathbf{w}) + (\mathbf{v}', \mathbf{w}') = (\mathbf{v} + \mathbf{v}', \mathbf{w} + \mathbf{w}').$$

Of course, one can similarly define the direct sum of several vector spaces. E.g. $\mathbb{K}^n = \mathbb{K}^1 \oplus \cdots \oplus \mathbb{K}^1$ (n times).

`Example 1.` Given a linear map $A : \mathcal{V} \to \mathcal{W}$, its **graph** is defined as a subspace in the direct sum $\mathcal{V} \oplus \mathcal{W}$:

$$\text{Graph } A = \{(\mathbf{v}, \mathbf{w}) \in \mathcal{V} \oplus \mathcal{W} : \ \mathbf{w} = A\mathbf{v}\}.$$

Affine Subspaces

Adding to all vectors of a subspace $\mathcal{W} \subset \mathcal{V}$ a fixed vector, $\mathbf{v} \in \mathcal{V}$, we construct an **affine subspace** (let us denote it $\mathbf{v} + \mathcal{W}$) parallel to $\mathcal{W}$:

$$\mathbf{v} + \mathcal{W} = \{\mathbf{v} + \mathbf{w} \mid \mathbf{w} \in \mathcal{W}\}.$$

In other words (Figure 24), $\mathbf{v}+\mathcal{W}$ is obtained from $\mathcal{W}$ by translation though the vector $\mathbf{v}$. When $\mathbf{v}$ lies in $\mathcal{W}$, the resulting affine subspace is linear as it coincides with $\mathcal{W}$. When $\mathbf{v}$ is not from $\mathcal{W}$, the resulting affine subspace has no common points with $\mathcal{W}$ (check it!), and is not linear (e.g. because $\mathbf{0} \notin \mathbf{v}+\mathcal{W}$). Yet it is often necessary to deal with such subspaces. For example, if $A : \mathcal{V} \to \mathcal{U}$ is a linear map, then for any point $\mathbf{u}$ in the range of A, the inverse image $A^{-1}(\mathbf{u})$ is an affine subspace in $\mathcal{V}$ parallel to the kernel of A. Indeed, if $\mathbf{v} \in A^{-1}(\mathbf{u})$, then all other vectors $\mathbf{v}'$ mapped by A to $\mathbf{u}$ are obtained by adding to $\mathbf{v}$ all vectors from the kernel: $A(\mathbf{v}') = \mathbf{u}$ if and only if $A(\mathbf{v}' - \mathbf{v}) = \mathbf{0}$.

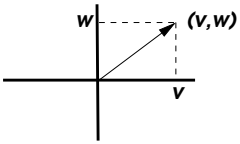

Figure 23. Direct sum

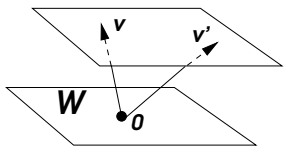

Figure 24. Affine subspace

The Dual Space

Let $\mathcal{V}^*$ denote the set of all **linear functions**[1] from a given $\mathbb{K}$-vector space $\mathcal{V}$ to the filed of scalars, i.e. functions $\mathbf{a} : \mathcal{V} \to \mathbb{K}$ such that

$$\mathbf{a}(\lambda\mathbf{u} + \mu\mathbf{v}) = \lambda\mathbf{a}(\mathbf{u}) + \mu\mathbf{a}(\mathbf{v})$$

for all $\mathbf{u}, \mathbf{v} \in \mathcal{V}$ and all $\lambda, \mu \in \mathbb{K}$. It is not hard to check that sums of linear functions are linear, and a scalar multiple of a linear function

[1] Theu are also called **linear forms** on $\mathcal{V}$.

is also linear. Moreover the set $\mathcal{V}^*$ equipped with such pointwise operations of addition of linear functions and their multiplication by scalars forms a $\mathbb{K}$-vector space. It is the **dual space** of $\mathcal{V}$.

`Example 2.` Coordinates $x_1, \dots, x_n$ on the standard coordinate space $\mathbb{K}^n$ are examples of linear forms on $\mathbb{K}^n$. Their linear combinations $a_1x_1 + \dots + a_nx_n$ with coefficients $a_1, \dots, a_n \in \mathbb{K}$ are also linear forms. In fact, as we will find out soon, all linear form on $\mathbb{K}^n$ are so described.

EXERCISES

128. Show that a map $A : \mathcal{V} \to \mathcal{W}$ is linear if and only if $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$ and $A(\lambda\mathbf{u}) = \lambda A\mathbf{u}$ for all $\mathbf{u}, \mathbf{v}$ in $\mathcal{V}$ and all $\lambda \in \mathbb{K}$.

129. Prove that rotation about the origin is a linear map from the plane to itself. ϟ

130. Show that every linear form on the coordinate (x, y)-plane is a linear combination of the coordinates: $\alpha x + \beta y$.

131. Establish an isomorphism between $\mathbb{K}^m \oplus \mathbb{K}^n$ and $\mathbb{K}^{m+n}$.

132. Show that $(\mathcal{V} \oplus \mathcal{W})^* = \mathcal{V}^* \oplus \mathcal{W}^*$. ϟ

133. Prove that for a linear map $A : \mathcal{V} \to \mathcal{W}$, $\operatorname{Ker} A = (\operatorname{Graph} A) \cap \mathcal{V}$.

134. Describe all affine subspaces in geometric 3D space. ✓

135.* Prove that the intersection of two affine subspaces, parallel to given linear ones, if non-empty, is an affine subspace parallel to the intersection of the given linear subspaces. ϟ

Further Examples of Vector Spaces

The axiomatic definition of vector spaces is *doubly* abstract: not only it neglects to specify the set $\mathcal{V}$ of vectors, but it does not even tell us anything explicit about the nature of the operations of addition of vectors and multiplication of vectors by scalars. To find various examples of vector spaces we should figure out which operations would be good candidates to satisfy the axioms (i–vii). It turns out that in the majority of useful examples, the operations are *pointwise addition of functions and multiplication of functions by scalars.*

`Example 3.` Let S be *any* set, and $\mathcal{V}$ be the set of *all* functions on S with values in $\mathbb{K}$. We will denote this set by $\mathbb{K}^S$. The sum and multiplication by scalars are defined on $\mathbb{K}^S$ as pointwise operations with functions. Namely, given two functions f, g and a scalar λ, the values of the sum $f + g$ and the product λf at a point $s \in S$ are

$$(f + g)(s) = f(s) + g(s), \ (\lambda f)(s) = \lambda(f(s)).$$

It is immediate to check that $\mathcal{V} = \mathbb{K}^S$ equipped with these operations satisfies the axioms (i–vii). Thus $\mathbb{K}^S$ is a $\mathbb{K}$-vector space.

`Example 3a.` Let S be the set of n elements $1, 2, ..., n$. Then the space $\mathbb{K}^S$ is the space $\mathbb{K}^n$ (e.g. $\mathbb{R}^S = \mathbb{R}^n$ and $\mathbb{C}^S = \mathbb{C}^n$) of **coordinate vectors**. Namely, each function on the set $\{1, \dots, n\}$ is specified by a string $\mathbf{x} = (x_1, ..., x_n)$ of its values, called **coordinates** or **components** of the coordinate vector. By tradition, the string is written as a **column**, and the pointwise operations with functions turn into termwise operations with the columns:

$$\lambda \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{bmatrix}, \quad \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}.$$

`Example 3b.` Let S be the set of all ordered pairs (i, j), where $i = 1, ..., m$, $j = 1, ..., n$. Then the vector space $\mathbb{K}^S$ is the space of $m \times n$-**matrices**. By tradition, a matrix is denoted by an upper-case letter, e.g. A, and is represented by a rectangular array whose **entry** at the intersection of the ith row and jth column is an element of $\mathbb{K}$ denoted by a_{ij} (the same lower-case letter with subscripts):

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}.$$

The pointwise operations with matrices as functions turn into elementwise addition of the arrays and their multiplication by scalars.

`Example 4.` A subspace of $\mathbb{K}$-vector space is an example of a $\mathbb{K}$-vector space.

`Example 4a.` An $m \times n$-matrix is called **square**, if $m = n$. A square matrix A is called **diagonal** (respectively, **upper-triangular**, or **lower-triangular**) if $a_{ij} = 0$ whenever $i \neq j$ (respectively, $i > j$, or $i < j$). Diagonal (respectively, upper-triangular, or lower-triangular) matrices form a subspace in the space of all $n \times n$-matrices, and therefore provide an example of a vector space.

`Example 4b.` The set of all polynomials (say, in one variable),[2] form a subspace in the space $\mathbb{R}^{\mathbb{R}}$ of all real-valued functions on the number line and therefore provide examples of real vector spaces. More generally, polynomials with coefficients in $\mathbb{K}$ (as well as such polynomials of degree not exceeding 7) form examples of $\mathbb{K}$-vector spaces.

`Example 5.` Let $\mathcal{V}$ be a $\mathbb{K}$-vector space. Consider the set $\mathcal{V}^S$ of all functions on a given set S with values in $\mathcal{V}$. Elements of $\mathcal{V}$ can be added

[2] As well as sets of all continuous, differentiable, 5 times continuously differentiable, infinitely differentiable, Riemann-integrable, measurable, etc. functions, introduced in mathematical analysis.

and multiplied by scalars. Respectively the *vector-valued* functions can be added and multiplied by scalars in the pointwise fashion. Thus, $\mathcal{V}^S$ is an example of a $\mathbb{K}$-vector space.

`Example 5a.` Linear combinations $\lambda A + \mu B$ of linear maps $A, B : \mathcal{V} \to \mathcal{W}$ are linear. Therefore all linear maps from $\mathcal{V}$ to $\mathcal{W}$ form a subspace in the space $\mathcal{W}^{\mathcal{V}}$ of all vector-valued functions $\mathcal{V} \to \mathcal{W}$. The vector space of linear maps from $\mathcal{V}$ to $\mathcal{W}$ is usually denoted by $\mathrm{Hom}(\mathcal{V}, \mathcal{W})$ (from the word *homomorphism*, synonimous in our context to the term *linear map*. By the very definition, $Hom(\mathcal{V}, \mathbb{K})$ is the space of all $\mathbb{K}$-linear forms $\mathcal{V} \to \mathbb{K}$, i.e. is the space $\mathcal{V}^*$ dual to $\mathcal{V}$.

The following formal construction indicates that *every vector space can be identified with a subspace in a space of functions with pointwise operations of addition and multiplication by scalars.*

`Example 5b.` Given a vector $\mathbf{v} \in \mathcal{V}$ and a linear function $f \in \mathcal{V}^*$, the value $f(\mathbf{v}) \in \mathbb{K}$ is defined. We can consider it not as a function f of $\mathbf{v}$, but as a function of f defined by $\mathbf{v}$. This way, to a vector $\mathbf{v}$ we associate the function $E_{\mathbf{v}} : \mathcal{V}^* \to \mathbb{K}$ defined by evaluating all linear functions $\mathcal{V} \to \mathbb{K}$ on the vector $\mathbf{v}$. The function $E_{\mathbf{v}}$ is linear, since $(\lambda f + \mu g)(\mathbf{v}) = \lambda f(\mathbf{v}) + \mu g(\mathbf{v})$. The linear function $E_{\mathbf{v}}$ is an element of the second dual space $(\mathcal{V}^*)^*$. The formula $f(\lambda \mathbf{v} + \mu \mathbf{w}) = \lambda f(\mathbf{v}) + \mu f(\mathbf{w})$, expressing linearity of linear functions, shows that $E_{\mathbf{v}}$ depends linearly on $\mathbf{v}$. Thus the **evaluation map** $E : \mathbf{v} \mapsto E_{\mathbf{v}}$ is a linear map $\mathcal{V} \to (\mathcal{V}^*)^*$. One can show that E ***is injective and thus provides an isomorphism between $\mathcal{V}$ and its range*** $E(\mathcal{V}) \subset (\mathcal{V}^*)^*$.

The previous result and examples suggest that vector spaces need not be described abstractly, and raises the suspicion that the axiomatic definition is misleading as it obscures the actual nature of vectors as functions subject to the pointwise algebraic operations. Here are however some examples where vectors do not come *naturally* as functions.

Perhaps, the most important example of this kind is given by geometric vectors, (as well as by forces and velocities in physics). It provides the opportunity to use geometric intuition in contexts unrelated to geometry. That is, one can "visualize" functions as geometric vectors. Furthermore, taking the field $\mathbb{Q}$ of rational numbers, or the field $\mathbb{Z}_2 = \{0, 1\}$ on the role of scalars, one can apply geometric intuition to number theory or computer science. Later we will have a chance to see how this works.

Another justification for introducing vector spaces abstractly is that this approach provides great flexibility in constructing new vector spaces from given ones. Such constructions (e.g. direct sums) are used regularly, and it would be very awkward to constantly express the resulting vector spaces as spaces of functions, even when the given spaces are expressed this way. Here is another example.

Quotient Spaces

The **quotient space** of a vector space $\mathcal{V}$ by a subspace $\mathcal{W}$ is defined as follows. Two vectors $\mathbf{v}$ and $\mathbf{v}'$ (Figure 24) are called *equivalent modulo* $\mathcal{W}$, if $\mathbf{v} - \mathbf{v}' \in \mathcal{W}$. This way, all vectors from $\mathcal{V}$ become partitioned into equivalence classes. These equivalence classes form the quotient vector space $\mathcal{V}/\mathcal{W}$.

In more detail, denote by $\pi : \mathcal{V} \to \mathcal{V}/\mathcal{W}$ the **canonical projection**, which assigns to a vector $\mathbf{v}$ its equivalence class modulo $\mathcal{W}$. This class can be symbolically written as $\mathbf{v} + \mathcal{W}$, a notation emphasizing that the class consists of all vectors obtained from $\mathbf{v}$ by adding arbitrary vectors from $\mathcal{W}$. Alternatively, one may think of $\mathbf{v} + \mathcal{W}$ as a "plane" obtained from $\mathcal{W}$ as *translation* by the vector $\mathbf{v}$. When $\mathbf{v} \in \mathcal{W}$, we have $\mathbf{v} + \mathcal{W} = \mathcal{W}$. When $\mathbf{v} \notin \mathcal{W}$, $\mathbf{v} + \mathcal{W}$ is not a linear subspace in $\mathcal{V}$. We will call it an **affine subspace** parallel to $\mathcal{W}$.

The set $\mathcal{V}/\mathcal{W}$ of all affine subspaces in $\mathcal{V}$ parallel to $\mathcal{W}$ is equipped with algebraic operations of addition and multiplication by scalars in such a way that the *canonical projection* $\pi : \mathcal{V} \to \mathcal{V}/\mathcal{W}$ *becomes a linear map.* In fact this condition leaves no choices, since it requires that for every $\mathbf{u}, \mathbf{v} \in \mathcal{V}$ and $\lambda, \mu \in \mathbb{K}$,

$$\lambda\pi(\mathbf{u}) + \mu\pi(\mathbf{v}) = \pi(\lambda\mathbf{u} + \mu\mathbf{v}).$$

In other words, the linear combination of given equivalence classes must coincide with the equivalence class containing the linear combination $\lambda\mathbf{u} + \mu\mathbf{v}$ of arbitrary representatives $\mathbf{u}, \mathbf{v}$ of these classes. It is important here that picking different representatives $\mathbf{u}'$ and $\mathbf{v}'$ will result in a new linear combination $\lambda\mathbf{u}' + \mu\mathbf{v}'$ which is however equivalent to the previous one. Indeed, the difference $\lambda(\mathbf{u} - \mathbf{u}') + \mu(\mathbf{v} - \mathbf{v}')$ lies in $\mathcal{W}$ since $\mathbf{u} - \mathbf{u}'$ and $\mathbf{v} - \mathbf{v}'$ do. Thus linear combinations in $\mathcal{V}/\mathcal{W}$ are well-defined.

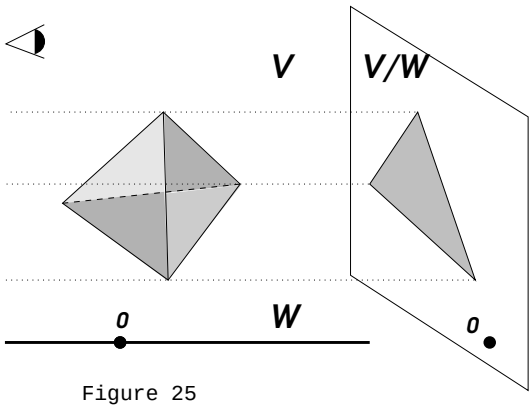

Figure 25

The construction of the quotient space is admittedly one of the most abstract ones so far. Here is a hint to how one could think of elements of $\mathcal{V}/\mathcal{W}$ and the projection π.

`Example 6.` Projecting $3D$ images to a 2-dimensional screen is described in geometry by the canonical projection π from the $3D$ space $\mathcal{V}$ to the plane $\mathcal{V}/\mathcal{W}$ of the screen along the line $\mathcal{W}$ of the eye sight (Figure 25).

`Example 7.` The direct sum $\mathcal{V} \oplus \mathcal{W}$ contains $\mathcal{V}$ and $\mathcal{W}$ as subspaces consisting of the pairs $(\mathbf{v}, \mathbf{0})$ and $(\mathbf{0}, \mathbf{w})$ respectively. The quotient of $\mathcal{V} \oplus \mathcal{W}$ by $\mathcal{W}$ is canonically identified with $\mathcal{V}$, because each pair $(\mathbf{v}, \mathbf{w})$ is equivalent modulo $\mathcal{W}$ to $(\mathbf{v}, \mathbf{0})$. Likewise, $(\mathcal{V} \oplus \mathcal{W})/\mathcal{V} = \mathcal{W}$.

`Example 8.` Let $\mathcal{V} = \mathbb{R}[x]$ be the space of polynomials with real coefficients, and $\mathcal{W}$ the subspace of polynomials divisible by $x^2 + 1$. Then the quotient space $\mathcal{V}/\mathcal{W}$ can be identified with the plane $\mathbb{C}$ of complex numbers, and the projection $\pi : \mathbb{R}[x] \to \mathbb{C}$ with the map $P \mapsto P(i)$ of evaluating a polynomial P at $x = i$. Indeed, polynomials P and P' are equivalent modulo $\mathcal{W}$ if and only if $P - P'$ is divisible by $x^2 + 1$, in which case $P(i) = P'(i)$. *Vice versa*, if $P(i) = P'(i)$, then $P(-i) = P'(-i)$ (since the polynomials are real), and hence $P - P'$ is divisible by $(x - i)(x + i) = x^2 + 1$.

For every linear map $A : \mathcal{V} \to \mathcal{V}'$, there is a canonical isomorphism $\tilde{A} : \mathcal{V}/\operatorname{Ker} A \to A(\mathcal{V})$ between the quotient by the kernel of A, and its range. Namely, $A\mathbf{u} = A\mathbf{v}$ if and only if $\mathbf{u} - \mathbf{v} \in \operatorname{Ker} A$, i.e. whenever $\mathbf{u}$ is equivalent to $\mathbf{v}$ modulo the kernel. Thus, *one can think of every linear map as the projection of the source space onto the range along the null space.* This is a manifestation of a general **homomorphism theorem** in algebra, which in the context of vector spaces can be formally stated this way:

`Theorem.` ***Every linear map*** $A : \mathcal{V} \to \mathcal{V}'$ ***is uniquely represented as the composition*** $A = i\tilde{A}\pi$ ***of the canonical projection*** $\pi : \mathcal{V} \to \mathcal{V}/\operatorname{Ker} A$ ***with the isomorphism*** $\tilde{A} : \mathcal{V}/\operatorname{Ker} A \to A(\mathcal{V})$ ***followed by the inclusion*** $i : A(\mathcal{V}) \subset \mathcal{V}'$***:***

$$\begin{array}{ccc} \mathcal{V} & \xrightarrow{A} & \mathcal{V}' \\ \pi \downarrow & & \cup\, i \\ \mathcal{V}/\operatorname{Ker} A & \xrightarrow[\tilde{A}]{\cong} & A(\mathcal{V}) \end{array} .$$

This result, although called a theorem, is merely a rephrasing of the definitions (of vector spaces, subspaces, quotient spaces, linear maps, isomorphisms, etc.) and is in this sense *tautological*,[3] void of new knowledge.

EXERCISES

136. Verify that $\mathbb{K}^S$ and $\mathcal{V}^S$ are vector spaces.

137. How many vectors are there in $\mathbb{Z}_p$-vector space $\mathbb{Z}_p^n$? $\operatorname{Hom}(\mathbb{Z}_p^n, \mathbb{Z}_p^m)$? ✓

[3]Dictionaries define **tautology** as "a representation of anything as the cause, condition, or consequence of itself."

138. How many vectors are there in the $\mathbb{Z}_p$-vector space of strictly upper triangular $n \times n$-matrices? ✓

139. Show that the map $f \mapsto \int_a^b f(x)\,dx$ defined by integration of (say) polynomial functions is a linear form $\mathbb{R}[x] \to \mathbb{R}$.

140. Find the kernel and the range of the differentiation map $D = \frac{d}{dx} : \mathbb{K}[x] \to \mathbb{K}[x]$, when (a) $\mathbb{K} = \mathbb{R}$, (b) $\mathbb{K} = \mathbb{Z}_p$. ✓

141.* Let $\mathcal{V} = \mathbb{R}[x]$, and $\mathcal{W} \subset \mathcal{V}$ be the subspace of all polynomials divisible by $x^2 - 1$. Establish an isomorphism between $\mathcal{V}/\mathcal{W}$ and $\mathbb{R}^{\{1,-1\}}$, the space of all functions from $\{1, -1\}$ to $\mathbb{R}$. ϟ

Bases and Dimension

Let $\mathcal{V}$ be a $\mathbb{K}$-vector space. A subset $V \subset \mathcal{V}$ (finite or infinite) is called a **basis** of $\mathcal{V}$ if every vector of $\mathcal{V}$ can be uniquely written as a (finite!) linear combination of vectors from V.

`Example 9.` Let $\mathbb{K}[x]$ denote the set of all polynomials in one indeterminate x with coefficients from $\mathbb{K}$. One can easily check, that it is a $\mathbb{K}$-vector space with respect to the operations of addition of polynomials and their multiplication by scalars. The monomilas x^k, $k = 0, 1, 2, \ldots$, form a basis in the space $\mathbb{K}[x]$ since every polynomial is uniquely written as a linear combination of monomials.

`Example 10.` In $\mathbb{K}^n$, every vector is uniquely written as the linear combination of **unit coordinate vectors** $\mathbf{e}_1, \ldots, \mathbf{e}_n$:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

Thus, the vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ form a basis. It is called the **standard basis** of $\mathbb{K}^n$.

The notion of basis has two aspects which can be considered separately.

Let $V \subset \mathcal{V}$ be *any* set of vectors. Linear combinations $\lambda_1\mathbf{v}_1 + \cdots + \lambda_k\mathbf{v}_k$, where the vectors $\mathbf{v}_i$ are taken from the subset V, and λ_i are arbitrary scalars, form a subspace in $\mathcal{V}$. (Indeed, sums and scalar multiples of linear combinations of vectors from V are also linear combinations of vectors from V.) This subspace is often denoted as $\operatorname{Span} V$. One says that the set V **spans** the subspace $\operatorname{Span} V$, or that $\operatorname{Span} V$ is **spanned** by V.

A set V of vectors is called **linearly independent**, if no vector from $\operatorname{Span} V$ can be represented as a linear combination of vectors from V in more than one way. To familiarize ourselves with this notion, let us give several reformulations of the definition. Here is one: no two *distinct* linear combinations of vectors from V are equal to each other. Yet another one: if two linear combinations of vectors from V *are* equal: $\alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k = \beta_1\mathbf{v}_1 + \cdot + \beta_k\mathbf{v}_k$, then their coefficients must be the same: $\alpha_1 = \beta_1, \ldots, \alpha_k = \beta_k$. Subtracting one linear combination from the other, we arrive at one more reformulation: if $\gamma_1\mathbf{v}_1 + \cdots + \gamma_k\mathbf{v}_k = \mathbf{0}$ for some vectors $\mathbf{v}_i \in V$, then necessarily $\gamma_1 = \cdots = \gamma_k = 0$. In other words, V is linearly independent, if the vector $\mathbf{0}$ can be represented as a linear combination of vectors from V only in the *trivial* fashion: $\mathbf{0} = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_k$. Equivalently, every **nontrivial linear combination** of vectors from V is not equal to zero: $\gamma_1\mathbf{v}_1 + \cdots + \gamma_k\mathbf{v}_k \neq \mathbf{0}$ whenever at least one of $\gamma_i \neq 0$.

Of course, a set $V \subset \mathcal{V}$ is called **linearly dependent** if it is not linearly independent. Yet, it is useful to have an affirmative reformulation: V is linearly dependent if and only if some *nontrivial* linear combination of vectors from V vanishes, i.e. $\gamma_1\mathbf{v}_1 + \cdots + \gamma_k\mathbf{v}_k = \mathbf{0}$, where at least one of the coefficients (say, γ_k) is non-zero. Dividing by this coefficient, and moving all other terms to the other side of the equality, we obtain one more reformulation: a set V is linearly dependent if one of its vectors can be represented as a linear combination of the others: $\mathbf{v}_k = -\gamma_k^{-1}\gamma_1\mathbf{v}_1 - \cdots - \gamma_k^{-1}\gamma_{k-1}\mathbf{v}_{k-1}$.[4] Obviously, every set containing the vector $\mathbf{0}$ is linearly dependent; every set containing two proportional vectors is linearly dependent; adding new vectors to a linearly dependent set leaves it linearly dependent.

Thus, a basis of $\mathcal{V}$ is a linearly independent set of vectors that spans the whole space.

In this course, we will be primarily concerned with **finite dimensional** vector spaces, i.e. spaces which can be spanned by finitely many vectors. If such a set of vectors is linearly dependent, then one of its vectors is a linear combination of the others. Removing this vector from the set, we obtain a smaller set that still spans $\mathcal{V}$. Continuing this way, we arrive at a finite linearly independent set that spans $\mathcal{V}$. Thus, a finite dimensional vector space has a basis, consisting of finitely many elements. The number of elements in a basis does not depend (as we will see shortly) on the choice of the basis. This number is called the **dimension** of the vector space $\mathcal{V}$

[4] It is essential here that division by all non-zero scalars is well-defined.

and is denoted $\dim \mathcal{V}$.

Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a basis of $\mathcal{V}$. Then every vector $\mathbf{x} \in \mathcal{V}$ is uniquely written as $\mathbf{x} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$. We call $(x_1, \ldots, x_n)$ **coordinates** of the vector $\mathbf{x}$ with respect to the basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$. For $\mathbf{y} = y_1\mathbf{v}_1 + \cdots + y_n\mathbf{v}_n \in \mathcal{V}$ and $\lambda \in \mathbb{K}$, we have:

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1)\mathbf{v}_1 + \cdots + (x_n + y_n)\mathbf{v}_n,$$
$$\lambda\mathbf{x} = (\lambda x_1)\mathbf{v}_1 + \cdots + (\lambda_n x_n)\mathbf{v}_n.$$

This means that operations of addition of vectors and multiplication by scalars are performed *coordinate-wise*. In other words, ***the map:***

$$\mathbb{K}^n \to \mathcal{V}: \quad x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n \mapsto x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$$

defines an isomorphism of the coordinate space $\mathbb{K}^n$ ***onto the vector space*** $\mathcal{V}$ ***with a basis*** $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$**.**

`Lemma.` ***A set of*** $n+1$ ***vectors in*** $\mathbb{K}^n$ ***is linearly dependent.***

`Proof.` Any two vectors in $\mathbb{K}^1$ are proportional and therefore linearly dependent. We intend to prove the lemma by deducing from this that any 3 vectors in $\mathbb{K}^2$ are linearly dependent, then deducing from this that any 4 vectors in $\mathbb{K}^3$ are linearly dependent, and so on. Thus we only need to prove that *if* every set of n vectors in $\mathbb{K}^{n-1}$ is linearly dependent *then* every set of $n+1$ vectors in $\mathbb{K}^n$ is linearly dependent too.[5]

To this end, consider $n+1$ column vectors $\mathbf{v}_1, ..., \mathbf{v}_{n+1}$ of size n each. If the bottom entry in each column is 0, then $\mathbf{v}_1, ..., \mathbf{v}_{n+1}$ are effectively $n-1$-columns. Hence some nontrivial linear combination of $\mathbf{v}_1, ..., \mathbf{v}_n$ is equal to $\mathbf{0}$ (by the induction hypothesis), and thus the whole set is linearly dependent. Now consider the case when at least one column has the bottom entry non-zero. Reordering the vectors we may assume that it is the column $\mathbf{v}_{n+1}$. Subtracting the column $\mathbf{v}_{n+1}$ with suitable coefficients $\alpha_1, ..., \alpha_n$ from $\mathbf{v}_1, ..., \mathbf{v}_n$ we can form n new columns $\mathbf{u}_1 = \mathbf{v}_1 - \alpha_1\mathbf{v}_{n+1}, ..., \mathbf{u}_n = \mathbf{v}_n - \alpha_n\mathbf{v}_{n+1}$ so that all of them have the bottom entries equal to zero. Thus $\mathbf{u}_1, ..., \mathbf{u}_n$ are effectively $n-1$-vectors and are therefore linearly dependent: $\beta_1\mathbf{u}_1 + ... + \beta_n\mathbf{u}_n = \mathbf{0}$ for some $\beta_1, ..., \beta_n$ not all equal to 0. Thus

$$\beta_1\mathbf{v}_1 + ... + \beta_n\mathbf{v}_n - (\alpha_1\beta_1 + ... + \alpha_n\beta_n)\mathbf{v}_{n+1} = \mathbf{0}.$$

[5]This way of reasoning is called **mathematical induction**. Put abstractly, it establishes a *sequence* P_n of propositions in two stages called respectively the **base** and **step** of induction: (i) P_1 is true; (ii) for all $n = 2, 3, 4, \ldots$, if P_{n-1} is true (the **induction hypothesis**) then P_n is true.

Here at least one of $\beta_i \neq 0$, and hence $\mathbf{v}_1, ..., \mathbf{v}_{n+1}$ are linearly dependent. □

`Corollaries.` (1) ***Any set of*** $m > n$ ***vectors in*** $\mathbb{K}^n$ ***is linearly dependent.***

(2) $\mathbb{K}^n$ ***and*** $\mathbb{K}^m$ ***are not isomorphic unless*** $n = m$***.***

(3) ***Every finite dimensional vector space is isomorphic to exactly one of the spaces*** $\mathbb{K}^n$***.***

(4) ***In a finite dimensional vector space, all bases have the same number of elements. In particular, dimension is well-defined.***

(5) ***Two finite dimensional vector spaces are isomorphic if and only if their dimensions are equal.***

Indeed, (1) is obvious because adding new vectors to a linearly dependent set leaves it linearly dependent. Since the standard basis in $\mathbb{K}^m$ consists of m linearly independent vectors, $\mathbb{K}^m$ cannot be isomorphic to $\mathbb{K}^n$ if $m > n$. This implies (2) and hence (3), because two spaces isomorphic to a third one are isomorphic to each other. Now (4) follows, since the choice of a basis of n elements establishes an isomorphism of the space with $\mathbb{K}^n$. Rephrasing (3) in terms of dimensions yields (5).

`Example 11.` Let $\mathcal{V}$ be a vector space of dimension n, and let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a basis of it. Then every vector $\mathbf{x} \in \mathcal{V}$ can be uniquely written as $\mathbf{x} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$. Here $x_1, \ldots, x_n$ can be considered as linear functions from $\mathcal{V}$ to $\mathbb{K}$. Namely, the function x_i takes the value 1 on the vector $\mathbf{v}_i$ and the value 0 on all $\mathbf{v}_j$ with $j \neq i$. Every linear function $f : \mathcal{V} \to \mathbb{K}$ takes on a vector $\mathbf{x}$ the value $f(\mathbf{x}) = x_1 f(\mathbf{v}_1) + \cdots + x_n f(\mathbf{v}_n)$. Therefore f is the linear combination of $x_1, \ldots, x_n$ with the coefficients $f(\mathbf{v}_1), \ldots, f(\mathbf{v}_n)$, i.e. $x_1, \ldots, x_n$ span the dual space $\mathcal{V}^*$. In fact, they are linearly independent, and thus form a basis of $\mathcal{V}^*$. Indeed, if a linear combination $\gamma_1 x_1 + \cdots + \gamma_n x_n$ coincides with the identically zero function, then its values γ_i on the vectors $\mathbf{v}_i$ must be all zeroes. We conclude that ***the dual space*** $\mathcal{V}^*$ ***has the same dimension*** n ***as*** $\mathcal{V}$ ***and is isomorphic to it.*** The basis $x_1, \ldots, x_n$ is called the **dual basis** of $\mathcal{V}^*$ with respect to the basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathcal{V}$.

`Remark.` Corollaries 3, 5, and Example 11 suggest that in a sense there is "only one" $\mathbb{K}$-vector space in each dimension $n = 0, 1, 2, \ldots$, namely $\mathbb{K}^n$. The role of this fact, which is literally true if the uniqueness is understood *up to isomorphism*, should not be overestimated.

An isomorphism $\mathbb{K}^n \to \mathcal{V}$ is determined by the choice of a basis in $\mathcal{V}$, and is therefore not unique. For example, the space of polynomials of degree $< n$ in one indeterminate x has dimension n and is isomorphic to $\mathbb{K}^n$. However, different isomorphisms may be useful for different purposes. In elementary algebra one would use the basis $1, x, x^2, \ldots, x^{n-1}$. In Calculus $1, x, x^2/2, \ldots, x^{n-1}/(n-1)!$ may be more common. In the theory of interpolation the basis of **Lagrange polynomials** is used:

$$L_i(x) = \frac{\prod_{j \neq i}(x - x_j)}{\prod_{j \neq i}(x_i - x_j)}.$$

Here $x_1, \ldots, x_n$ are given distinct points on the number line, and $L_i(x_j) = 0$ for $j \neq i$ and $L_i(x_i) = 1$. The theory of orthogonal polynomials leads to many other important bases, e.g. those formed by **Chebyshev polynomials**[6] T_k, or **Hermite polynomials** H_k:

$$T_k(x) = \mathbf{cos}\left(k\,\mathbf{cos}^{-1}(x)\right), \qquad H_k(x) = e^{x^2}\frac{d^k}{dx^k}e^{-x^2}.$$

There is no preferred basis in an n-dimensional vector space $\mathcal{V}$ (and hence no preferred isomorphism between $\mathcal{V}$ and $\mathbb{K}^n$).

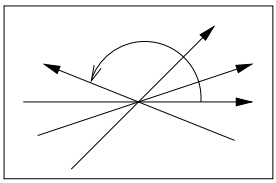

Figure 26. Subspaces

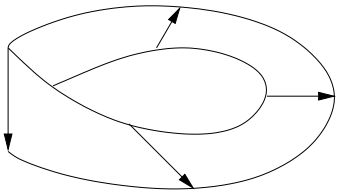

Figure 27. The Mobius band

The lack of the preferred isomorphism becomes really important when *continuous families* of vector spaces get involved. For instance, consider on the plane $\mathbb{R}^2$, *all* subspaces of dimension 1 (Figure 26). When subspaces rotate, one can pick a basis vector in each of them, which would vary continuously, but when the angle of rotation approaches π, the *direction* of the vector disagrees with the initial direction of the same line. In fact it is impossible to choose bases in all the lines in a continuous fashion. The reason is shown on Figure 27: The surface formed by the continuous family of 1-dimensional

[6] After Pafnuty **Chebyshev** (1821– 1984).

subspaces in $\mathbb{R}^2$ has the topology of a **Möbius band** (rather than a cylinder). The Möbius band is a first example of nontrivial *vector bundles*. Vector bundles are studied in Homotopic Topology. It turns out that among all k-dimensional vector bundles (i.e. continuous families of k-dimensional vector spaces) the most complicated are the bundles formed by *all* k-dimensional subspaces in the coordinate space of dimension $n \gg k$.

Likewise, though Example 11 shows that a finite dimensional vector space $\mathcal{V}$ is isomorphic to its dual $\mathcal{V}^*$, the isomorphism depends on the choice of a basis in $\mathcal{V}$, and the same is true about $\mathcal{V}^*$ and $\mathcal{V}^{**}$, the space of linear functions $\mathcal{V}^* \to \mathbb{K}$. However, the last of the exercises below shows that a finite-dimensional vector space $\mathcal{V}$ is **canonically isomorphic** to its second dual $\mathcal{V}^{**}$, i.e. that there *is* a preferred isomorphism between $\mathcal{V}$ and $\mathcal{V}^{**}$. It is defined by interpreting the value of a linear form on a vector as a rule that to a given vector associates a function of the linear form.

EXERCISES

142. Describe all bases in $\mathbb{K}^1$.

143. Prove that the set consisting of one vector is linearly dependent if and only if the vector is $\mathbf{0}$.

144. Prove that a subset of a linearly independent set of vectors is linearly independent.

145. Prove that a set of vectors containing a linearly dependent subset is linearly dependent.

146. Prove that the set of four polynomials $x^2, (x-1)^2, (x-2)^2, (x-3)^2$ is linearly dependent, but any proper subset of it is linearly independent.

147. Find a basis of the subspace in $\mathbb{R}^3$ given by the equation $x_1+x_2+x_3 = 0$.

148. The same, for the subspace in $\mathbb{C}^4$ given by the equation $x_1 + x_2 + x_3 + x_4 = 0$.

149. Find a basis in and the dimension of the subspace in $\mathbb{K}^4$ given by two equations: $x_1 + x_2 + x_3 = 0$ and $x_2 + x_3 + x_4 = 0$. ✓

150. Prove that $\mathbb{K}[x]$ is infinite-dimensional.

151.* In the space of polynomials of degree $< n$, express the basis $x^k, k = 0, 1, \dots, n-1$ of monomials in terms of the basis $L_i, i = 1, \dots, n$, of Lagrange polynomials. ✓

152. In $\mathbb{R}[x]$, find coordinates of the Chebyshev polynomials T_4 in the basis of monomials x^k, $k = 0, 1, 2, \dots$ ϟ

153. Given a vector space $\mathcal{V}$, associate to a vector $\mathbf{v} \in \mathcal{V}$ the function

$E_{\mathbf{v}} : \mathcal{V}^* \to \mathbb{K}$ defined by evaluation of linear forms $\mathbf{a} \in \mathcal{V}^*$ on the vector $\mathbf{v}$: $E_{\mathbf{v}}(\mathbf{a}) := \mathbf{a}(\mathbf{v})$. Prove that $E_{\mathbf{v}} \in \mathcal{V}^{**}$, and that the map $E : \mathcal{V} \to \mathcal{V}^{**} : \mathbf{v} \mapsto E_{\mathbf{v}}$ is linear.

154.* For a finite-dimensional vector space $\mathcal{V}$, show that the map $E : \mathcal{V} \to \mathcal{V}^{**}$, defined by evaluation of linear forms on vectors, is an isomorphism. ϟ

155. Find the dimensions of spaces of: (a) diagonal $n \times n$-matrices; (b) upper-triangular $n \times n$-matrices. ✓

156. Find the dimension of the subspace in $\mathbb{R}^{\mathbb{R}}$ spanned by functions $\mathbf{cos}(x + \theta_1), \dots, \mathbf{cos}(x + \theta_n)$, where $\theta_1, \dots, \theta_n$ are given distinct angles. ✓

157. Let $\mathcal{W} \subset \mathcal{V}$ be any subset of $\mathcal{V}$. Define $\mathcal{W}^\perp \subset \mathcal{V}^*$ as the set of all those linear functions which vanish on $\mathcal{W}$. Prove that $\mathcal{W}^\perp$ is a subspace of $\mathcal{V}^*$. (It is called the **annihilator** of $\mathcal{W}$.)

158. Let $\mathcal{W} \subset \mathcal{V}$ be a subspace. Establish a canonical isomorphism between the dual space $(\mathcal{V}/\mathcal{W})^*$ and the annihilator $\mathcal{W}^\perp \subset \mathcal{V}^*$. ϟ

159. Show that a field $\mathbb{K}$ can be considered as an $\mathbb{F}$-vector space over any subfield $\mathbb{F} \subset \mathbb{K}$.

160. Prove that the number of elements in a finite field is a power of its characteristic. ϟ

2 Matrices

Matrices are rectangular arrays of numbers. An $m \times n$-matrix

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \dots & a_{ij} & \dots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$$

has m rows and n columns. The **matrix entry** a_{ij} is positioned in row i and column j.

In linear algebra, matrices are found all over the place. Yet, the main point of this section is that matrices *per se* are not objects of linear algebra. Namely, the same matrix can represent different mathematical objects, and may behave differently depending on what kind of object is meant.

More precisely, various geometric objects, such as vectors, linear functions, quadratic forms, linear maps, etc., when expressed in coordinates, are represented by matrices. It is one of our goals in this book to demonstrate advantages of the *geometric* (i.e. "conceptual," coordinate-less) way of thinking over the *algebraic* (i.e. "computational,", coordinate-wise) attitude. So, we begin here with going as far as possible in the oppositite direction by examining the matrix, coordinate expressions of all geometric objects of our interest. Whenever suitable, we derive properties of operations with matrices from the properties of the geometric objects they represent (and not the other way around).

Vectors and Linear Functions

Let $\mathcal{V}$ be a finite dimensional vector space. Picking a basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ identifies each vector $\mathbf{x} = x_1\mathbf{e}_1 + \dots + x_n\mathbf{e}_n$ with the $n \times 1$-matrix, the *column* of its coordinates $x_1, \dots, x_n$. As we saw in the previous section, the results of vector operations: the multiplication by scalars $\lambda\mathbf{x}$ and the sum $\mathbf{x} + \mathbf{y}$, are expressed by componentwise operations with the columns:

$$\lambda \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{bmatrix}, \quad \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}.$$

A **linear function** (or **linear form**) $\mathbf{a} : \mathcal{V} \to \mathbb{K}$ is determined by its values on the basis vectors:

$$\mathbf{a}(\mathbf{x}) = x_1\mathbf{a}(\mathbf{e}_1) + \dots x_n\mathbf{a}(\mathbf{e}_n) = a_1x_1 + \dots + a_nx_n, \quad a_i := \mathbf{a}(\mathbf{e}_i).$$

Conversely, each function of the form $a_1x_1 + \cdots + a_nx_n$ is linear (for it is a linear combination of the coordinate functions x_i, which are linear). Linear functions are traditionally represented by $1 \times n$-matrices, the *rows* of their coefficients: $[a_1, \ldots, a_n]$.

Linear combinations $\lambda\mathbf{a} + \mu\mathbf{b}$ of linear functions are linear functions. Their coefficients are expressed by linear combinations of the rows:

$$\lambda[a_1, \ldots, a_n] + \mu[b_1, \ldots, b_n] = [\lambda a_1 + \mu b_1, \ldots, \lambda a_n + \mu b_n].$$

The operation of **evaluation**, i.e. taking the value $\mathbf{a}(\mathbf{x})$ of a linear function on a vector, is expressed in coordinates by the simplest instance of **matrix product**, namely the product of a $1 \times n$ row with an $n \times 1$ column:

$$\mathbf{a}(\mathbf{x}) = [a_1, \ldots, a_n] \begin{bmatrix} x_1 \\ \ldots \\ x_n \end{bmatrix} = a_1x_1 + \cdots + a_nx_n.$$

Note that by the very definition of linear functions,

$$\mathbf{a}(\lambda\mathbf{x} + \mu\mathbf{y}) = \lambda\mathbf{a}(\mathbf{x}) + \mu\mathbf{a}(\mathbf{y}) \text{ for all vectors } \mathbf{x}, \mathbf{y} \text{ and scalars } \lambda, \mu.$$

This can be viewed as a manifestation of the **distributive law** for matrix product.

EXERCISES

161. Are the functions $3x$, x^3, $x{+}1$, 0, $\mathbf{sin}\, x$, $(1{+}x)^2{-}(1{-}x)^2$, $\mathbf{tan}\,\mathbf{arctan}\, x$, $\mathbf{arctan}\,\mathbf{tan}\, x$ linear? ✓

162. Verify the *linearity* property of the coordinate functions x_i on $\mathbb{K}^n$.

Linear Maps

Let $A : \mathcal{V} \to \mathcal{W}$ be a linear map between two finite dimensional vector spaces. Picking a basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ in $\mathcal{V}$ and a basis $\{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$ in $\mathcal{W}$, one can express $A\mathbf{e}_j$ as linear combinations of $\mathbf{f}_i$:

$$A\mathbf{e}_j = a_{1j}\mathbf{f}_1 + \cdots + a_{mj}\mathbf{f}_m, \quad j = 1, \ldots, n.$$

The whole map A is uniquely determined by the coefficients a_{ij}; namely, if $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$, then by the very definition of linearity, we have:

$$A\mathbf{x} = x_1A\mathbf{e}_1 + \cdots + x_nA\mathbf{e}_n = x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix}.$$

In other words, the map $\mathbf{y} = A\mathbf{x}$ is given by m linear functions

$$y_i = \mathbf{a}_i(\mathbf{x}) = a_{i1}x_1 + \cdots + a_{in}x_n, \quad i = 1, \ldots, m.$$

Whereas each y_i is given by the product of a row with a column, the whole linear map can be described as the product of the $m \times n$-matrix $A = [a_{ij}]$ with the column:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = A\mathbf{x}.$$

Thus, the rows of the matrix A represent the m linear functions $\mathbf{a}_1, \ldots, \mathbf{a}_m$, the columns represent the images $A\mathbf{e}_1, \ldots, A\mathbf{e}_n$ in $\mathcal{W}$ of the basis vectors of $\mathcal{V}$, and the matrix entries $a_{ij} = \mathbf{a}_i(\mathbf{e}_j)$ represent the values of the linear functions on the basis vectors.

Conversely, in coordinates, the rows of any $m \times n$-matrix determine m linear functions $y_i = \mathbf{a}_i(\mathbf{x})$, which altogether determine a map from $\mathcal{V} = \mathbb{K}^n$ to $\mathbb{K}^m = \mathcal{W}$, which is automatically linear (since the functions $\mathbf{a}_i$ are).

EXERCISES

163. Let $\mathbf{v} \in \mathbb{K}^m$, and let $\mathbf{a} : \mathbb{K}^n \to \mathbb{K}$ be a linear function. Define a linear map $E : \mathbb{K}^n \to \mathbb{K}^m$ by $E(\mathbf{x}) = \mathbf{a}(\mathbf{x})\mathbf{v}$, and compute the matrix of E. ✓

164. Write down the matrix of rotation through the angle θ in $\mathbb{R}^2$. ✓

165. Let $\sum_{j=1}^{n} a_{ij}x_j = b_i, \quad i = 1, \ldots, m$, be a system of m **linear equations** in n unknowns $(x_1, \ldots, x_n)$. Show that it can be written in the matrix form $A\mathbf{x} = \mathbf{b}$, where A is a linear map from $\mathbb{K}^n$ to $\mathbb{K}^m$.

Composition

Given linear maps $B : \mathcal{U} \to \mathcal{V}$ and $A : \mathcal{V} \to \mathcal{W}$, one can form their **composition** $C : \mathcal{U} \to \mathcal{W}$ by substituting $\mathbf{v} = B\mathbf{u}$ into $A\mathbf{v}$: $C\mathbf{u} := A(B(\mathbf{u}))$. When the spaces are finite dimensional, $\dim \mathcal{U} = l$, $\dim \mathcal{V} = n$, $\dim \mathcal{W} = m$, one can pick their bases (thereby identifying the spaces with $\mathbb{K}^l, \mathbb{K}^n$, and $\mathbb{K}^m$ respectively), and express the operation in terms of the corresponding matrices. This leads to the general notion of **matrix product**, $C = AB$, defined whenever the number of columns of A coincides with the number of rows of B:

$$\begin{bmatrix} c_{11} & \ldots & c_{1l} \\ \ldots & c_{ij} & \ldots \\ c_{m1} & \ldots & c_{ml} \end{bmatrix} = \begin{bmatrix} a_{11} & \ldots . & a_{1n} \\ \ldots & & \ldots \\ a_{m1} & \ldots . & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & \ldots & b_{1l} \\ \ldots & & \ldots \\ \ldots & & \ldots \\ b_{n1} & \ldots & b_{nl} \end{bmatrix}.$$

By definition of the composition, the entry c_{ij} located at the intersection of the ith row and jth column of C is the value of the linear function $\mathbf{a}_i$ on the image $B\mathbf{e}_j$ in $\mathbb{K}^n$ of the standard basis vector $\mathbf{e}_j$ from $\mathbb{K}^l$. Since $\mathbf{a}_i$ and $B\mathbf{e}_j$ are represented by the ith row and jth column respectively of the matrices A and B, we find:

$$c_{ij} = [a_{i1}, \dots, a_{in}] \begin{bmatrix} b_{1j} \\ \dots \\ b_{nj} \end{bmatrix} = a_{i1}b_{1j} + \dots + a_{in}b_{nj}.$$

In other words, c_{ij} is the product of the ith row of A with the jth column of B.

Based on this formula, it is not hard to verify that the matrix product is **associative**, i.e. $(AB)C = A(BC)$, and satisfies the *left* and *right* distributive laws: $P(\lambda Q + \mu R) = \lambda PQ + \mu PR$ and $(\lambda X + \mu Y)Z = \lambda XZ + \mu YZ$, whenever the sizes of the matrices allow to form the expressions. However, our point is that there is no point in making such verifications. The operations with matrices encode in the coordinate form meaningful operations with linear maps, and the properties of matrices simply reflect those of the maps. For instance, the matrix product is associative because composition of *arbitrary* (and not only linear) maps is associative.

EXERCISES

166. Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & -1 & 1 \end{bmatrix},\ B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix},\ C = \begin{bmatrix} 1 & 2 \\ -2 & 1 \\ 0 & -1 \end{bmatrix}.$$

Compute those of the products $ABC, BAC, BCA, CBA, CAB, ACB$ which are defined. ✓

167. Let X, Y, Z, W be four sets, and $h : X \to Y$, $g : Y \to Z$, and $f : Z \to W$ be three functions. Show that the compositions $f \circ (g \circ h)$ and $(f \circ g) \circ h$ coincide. ϟ

168. Prove that products of upper triangular matrices are upper triangular and products of lower triangular matrices are lower triangular.

169.* Prove that a matrix of a linear transformation from $\mathbb{K}^n$ to itself is upper triangular if and only if the linear transformation maps each of the subspaces $Span(\mathbf{e}_1) \subset Span(\mathbf{e}_1, \mathbf{e}_2) \subset Span(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) \subset ...$ to itself. Derive from this that products of upper triangular matrices is upper triangular.

Changes of Coordinates

Let us examine here how the components of vectors, the coefficients of linear forms, and the matrices of linear maps are transformed under linear changes of coordinates.

Suppose that a finite dimensional vector space $\mathcal{V}$ is identified with $\mathbb{K}^n$ by the choice of a basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, and then identified with it in another way by the choice of a new basis, $\{\mathbf{e}'_1, \dots, \mathbf{e}'_n\}$. We can express the vectors of the new basis as linear combinations of the old basis vectors, and substitute these expressions into two coordinate representations of the same vector in the old and new coordinate systems:

$$\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n = x'_1\mathbf{e}'_1 + \cdots + x'_n\mathbf{e}'_n.$$

This will result in expressing the old coordinate functions x_i in terms of the new coordinate functions x'_j:

$$\begin{bmatrix} x_1 \\ \dots \\ x_n \end{bmatrix} = \begin{bmatrix} c_{11} & \dots & c_{1n} \\ \dots & & \dots \\ c_{n1} & \dots & c_{nn} \end{bmatrix} \begin{bmatrix} x'_1 \\ \dots \\ x'_n \end{bmatrix}.$$

In matrix notation, this can be written as $\mathbf{x} = C\mathbf{x}'$ where C is a **square matrix** of size n.

Conversely, $\mathbf{x}'$ can be expressed by linear functions of $\mathbf{x}$. In other words, there exists a square matrix D such that the substitutions $\mathbf{x}' = D\mathbf{x}$ and $\mathbf{x} = C\mathbf{x}'$ are inverse to each other, i.e. $\mathbf{x} = CD\mathbf{x}$ and $\mathbf{x}' = DC\mathbf{x}'$ for all $\mathbf{x}$ and $\mathbf{x}'$. It is immediate to see that this happens exactly when the matrices C and D satisfy $CD = I = DC$ where I is the **identity matrix** of size n:

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & & & \dots \\ 0 & \dots & 0 & 1 \end{bmatrix}.$$

When this happens, the square matrices C and D are called **inverse** to each other, and one writes: $D = C^{-1}$, $C = D^{-1}$.

To reiterate: the rows of C express old coordinates x_i as linear functions of new coordinates, x'_j. The columns of C represent, in the old coordinate system, the vectors which in the new coordinate system serve as standard basis vectors: $\mathbf{e}'_j = c_{1j}\mathbf{e}_1 + \cdots + c_{nj}\mathbf{e}_n$. The matrix C is often called the **transition matrix** between the coordinate systems.

In spite of apparent simplicity of this notion, it is easy to get lost here, for a change of coordinates is easy to confuse with a linear map from the space $\mathbb{K}^n$ *to itself*. For linear maps from a vector space to itself, we will reserve the term **linear transformation**. Thus, the same square matrix C defines the linear transformation $\mathbf{x}' = C\mathbf{x}$ which associates to a vector $\mathbf{x} \in \mathbb{K}^n$ a new vector $\mathbf{x}'$ written in *the same* coordinate system. The **inverse transformation**, when exists, is given by the formula $\mathbf{x} = C^{-1}\mathbf{x}'$.

Returning to changes of coordinates, examine how they affect linear functions and maps.

Making the change of variables $\mathbf{x} = C\mathbf{x}'$ in a linear function $\mathbf{a}$ results in the values $\mathbf{a}(\mathbf{x})$ of this function being expressed as $\mathbf{a}'(\mathbf{x}')$ in terms of new coordinates of the same vectors. Using the matrix product notation, we find: $\mathbf{a}'\mathbf{x}' = \mathbf{a}\mathbf{x} = \mathbf{a}(C\mathbf{x}') = (\mathbf{a}C)\mathbf{x}'$. Thus, coordinates of vectors and coefficients of linear functions are transformed differently:

$$\mathbf{x} = C\mathbf{x}', \quad \text{or} \quad \mathbf{x}' = C^{-1}\mathbf{x}, \quad \text{but} \quad \mathbf{a}' = \mathbf{a}C, \quad \text{or} \quad \mathbf{a} = \mathbf{a}'C^{-1}.$$

Next, let $\mathbf{y} = A\mathbf{x}$ be a linear map from $\mathbb{K}^n$ to $\mathbb{K}^m$, and let $\mathbf{x} = C\mathbf{x}'$ and $\mathbf{y} = D\mathbf{y}'$ be changes of coordinates in $\mathbb{K}^n$ and $\mathbb{K}^m$ respectively. Then in new coordinates the same linear map is given by the new formula $\mathbf{y}' = A'\mathbf{x}'$. We compute A' in terms of A, C and D: $D\mathbf{y}' = \mathbf{y} = A\mathbf{x} = AC\mathbf{x}'$, i.e. $\mathbf{y}' = D^{-1}AC\mathbf{x}'$, and hence

$$A' = D^{-1}AC.$$

In particular, suppose that $\mathbf{x} \mapsto A\mathbf{x}$ is a linear transformation on $\mathbb{K}^n$ (i.e. — we remind — a linear map from $\mathbb{K}^n$ to itself), and a change of coordinates $\mathbf{x} = C\mathbf{x}'$ is made. Then in new coordinates the same linear transformation is $\mathbf{x}' \mapsto A'\mathbf{x}'$, where

$$A' = C^{-1}AC,$$

i.e. in the previous rule we need to take $D = C$. This is because the same change applies to both: the input vector $\mathbf{x}$ and its image $A\mathbf{x}$. The operation $A \mapsto C^{-1}AC$ over a *square* matrix A is often called the **similarity transformation** by the invertible matrix C.

EXERCISES

170. For an identity matrix I, prove $AI = A$ and $IB = B$ for all allowed sizes of A and B.

171. For a square matrix A, define its **powers** A^k for $k > 0$ as $A \cdots A$ (k times), for $k = 0$ as I, and for $k < 0$ as $A^{-1} \cdots A^{-1}$ (k times), assuming A invertible. Prove that A that $A^k A^l = A^{k+l}$ for all integer k, l.

172.* Compute $\begin{bmatrix} \cos 19^\circ & -\sin 19^\circ \\ \sin 19^\circ & \cos 19^\circ \end{bmatrix}^{19}$. ✓

173. Compute powers $A^k, k = 0, 1, 2, \ldots$, of the square matrix A all of whose entries are zeroes, except that $a_{i,i+1} = 1$ for all i. ✓

174.* Let N be a linear transformation from $\mathbb{K}^n$ to itself, determined by $N\mathbf{e}_1 = 0, N\mathbf{e}_2 = \mathbf{e}_1, N\mathbf{e}_3 = \mathbf{e}_2, \ldots, N\mathbf{e}_n = \mathbf{e}_{n-1}$. Describe the maps N^k, $k = 1, 2, \ldots$, and find their matrices. ✓

175. For which sizes of matrices A and B, both products AB and BA: (a) are defined, (b) have the same size? ✓

176.* Give examples of matrices A and B which do not *commute*, i.e. $AB \neq BA$, even though both products are defined and have the same size. ϟ

177. When does $(A + B)^2 = A^2 + 2AB + B^2$? ✓

178. Which diagonal matrices are invertible?

179. Prove that an inverse of a given matrix is unique when exists. ϟ

180. Let A, B be invertible $n \times n$-matrices. Prove that AB is also invertible, and $(AB)^{-1} = B^{-1}A^{-1}$. ϟ

181.* If AB is invertible, does it imply that A, B, and BA are invertible? ϟ

182.* Give an example of matrices A and B such that $AB = I$, but $BA \neq I$.

183. Let $A : \mathbb{K}^n \to \mathbb{K}^m$ be an isomorphism. (Thus, $m = n$, but we consider $\mathbb{K}^n$ and $\mathbb{K}^m$ as two different copies of the coordinate space.) Prove that after suitable changes of coordinates in $\mathbb{K}^n$ and $\mathbb{K}^m$, the matrix of this transformation becomes the identity matrix I. ϟ

Transposition

Let $A : \mathcal{V} \to \mathcal{W}$ be a linear map, and $\mathbf{a} : \mathcal{W} \to \mathbb{K}$ a linear function. Composing the linear function with the map, we obtain a linear function on $\mathcal{V}$. This construction defines a map from $\mathcal{W}^*$, the space of linear functions on $\mathcal{W}$, to $\mathcal{V}^*$. The map is called **dual** (or **adjoint**, or **transposed**) to A, and denoted A^t. Thus $A^t : \mathcal{W}^* \to \mathcal{V}^*$ is defined by the formula

$$(A^t\mathbf{a})(\mathbf{v}) := \mathbf{a}(A\mathbf{v}).$$

It is a linear map. Indeed, if $\mathbf{c} = \lambda\mathbf{a} + \mu\mathbf{b}$, then $\mathbf{c}(A\mathbf{v}) = \lambda\mathbf{a}(A\mathbf{v}) + \mu\mathbf{b}(A\mathbf{v})$ for all $\mathbf{v}$, and hence $A^t\mathbf{c} = \lambda A^t\mathbf{a} + \mu A^t\mathbf{b}$. Note that A^t acts in the direction "opposite to A".

In coordinates, if $\mathbf{b} = A^t\mathbf{a}$, then

$$[b_1, \dots, b_n] = [a_1, \dots, a_m] \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \dots & a_{ij} & \dots \\ a_{m1} & \dots & a_{mn} \end{bmatrix},$$

that is, the row of the coefficients of the composite linear function $\mathbf{b}(\mathbf{x}) = \mathbf{a}(A\mathbf{x})$ is obtained from the row of the coeffiients of $\mathbf{a}$ by the multiplication on the right by the matrix A. This is a special case of the matrix expression for composition of linear maps. If however, we think of elements of $\mathcal{W}^*$ as "vectors," i.e. represent them by *columns* of their coordinates, then we have no other choice but to flip all the rows into columns here:

$$\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{m1} \\ \vdots & a_{ji} & \vdots \\ a_{1n} & \dots & a_{nm} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}.$$

Note that the matrix $[a_{ji}]$ is not A but is the matrix A^t **transposed** to A. By definition, this means that it is an $n \times m$-matrix, whose entry a^t_{ij} on the intersection of row i and column j coincides with the entry a_{ji} of the $m \times n$-matrix A, located at the intersection of the row j and column i. Thus, the equality can be written as $\mathbf{b}^t = A^t\mathbf{a}^t$.

Given an $m \times n$-matrix A and $n \times l$-matrix B, we have

$$(AB)^t = B^tA^t.$$

One can check this matrix identity directly, but in fact there is no need to do so. Indeed, given three vector spaces and two linear maps: $B : \mathcal{U} \to \mathcal{V}$, and $A : \mathcal{V} \to \mathcal{W}$, the adjoint map to their composition $AB : \mathcal{U} \to \mathcal{W}$ acts on linear forms $\mathcal{W} \to \mathbb{K}$ as the composition: $A^t : \mathcal{W}^* \to \mathcal{V}^*$, followed by $B^t : \mathcal{V}^* \to \mathcal{U}^*$. Indeed, for all $\mathbf{a} \in \mathcal{W}^*$ and all $\mathbf{u} \in \mathcal{U}$,

$$((AB)^t\mathbf{a})(\mathbf{u}) := \mathbf{a}(AB\mathbf{u}) = (A^t\mathbf{a})(B\mathbf{u}) = (B^t(A^t\mathbf{a}))(\mathbf{u}).$$

EXERCISES

184. Let A be the embedding $\mathcal{V} \subset \mathcal{W}$ of $\mathcal{V}$ as a linear subspace into $\mathcal{W}$. Describe the adjoint map $A^t : \mathcal{W}^* \to \mathcal{V}^*$. ✓

185. On the plane $z = 2x + 3y$ in the (x, y, z)-space, use (x, y) as a coordinate system, and express the restrictions of linear functions $\alpha x + \beta y + \gamma z$ to this plane in these coordinates. Can you identify in this computation the two transposed matrices A and A^t introduced in the previous exercise?

Bilinear Forms

The **dot-product** of two coordinate vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is defined by the formula

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + \cdots + x_n y_n.$$

It is an example of a **bilinear form**, i.e. a function of an ordered pair $(\mathbf{x}, \mathbf{y})$ of vector variables, which is a linear function of each of them.

In general, a **bilinear form** is a function $B : \mathcal{V} \times \mathcal{W} \to \mathbb{K}$, which to each ordered pair $(\mathbf{v}, \mathbf{w})$, where $\mathbf{v} \in \mathcal{V}$ and $\mathbf{w} \in \mathcal{W}$, associates a scalar $B(\mathbf{v}, \mathbf{w})$ in such a way that for a fixed $\mathbf{w}$ this is a linear function of $\mathbf{v}$, and for a fixed $\mathbf{v}$ a linear function of $\mathbf{w}$:

$$\begin{aligned} B(\lambda \mathbf{v} + \lambda' \mathbf{v}',\ \mathbf{w}) &= \lambda B(\mathbf{v}, \mathbf{w}) + \lambda' B(\mathbf{v}', \mathbf{w}), \\ B(\mathbf{v}, \lambda \mathbf{w} + \lambda' \mathbf{w}') &= \lambda B(\mathbf{v}, \mathbf{w}) + \lambda' B(\mathbf{v}, \mathbf{w}') \\ \text{for all } \mathbf{v}, \mathbf{v}' \in \mathcal{V},\ \mathbf{w}, \mathbf{w}' \in \mathcal{W},\ &\text{and } \lambda, \lambda' \in \mathbb{K}. \end{aligned}$$

Assuming that $\mathcal{V}$ and $\mathcal{W}$ are finite dimensional vector spaces, one can pick bases $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ in $\mathcal{V}$ and $\{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ in $\mathcal{W}$, and writing $\mathbf{v} = x_1 \mathbf{e}_1 + \cdots + x_m \mathbf{e}_m$, $\mathbf{w} = y_1 \mathbf{f}_1 + \cdots + y_n \mathbf{f}_n$, find:

$$B(\mathbf{v}, \mathbf{w}) = \sum_{i=1}^{m} x_i B(\mathbf{e}_i, \mathbf{w}) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_i B(\mathbf{e}_i, \mathbf{f}_j) y_j.$$

Thus, a bilinear form is uniquely determined by the $m \times n$-matrix of the coefficients $B(\mathbf{e}_i, \mathbf{f}_j)$. Conversely, every $m \times n$-matrix $B = [b_{ij}]$ defines a bilinear form of two coordinate vectors, $\mathbf{x} \in \mathbb{K}^m$ and $\mathbf{y} \in \mathbb{K}^n$, by the formula:

$$B(\mathbf{x}, \mathbf{y}) := \mathbf{x}^t B \mathbf{y} = \sum_{i=1}^{m} \sum_{j=1}^{n} x_i b_{ij} y_j.$$

It is easy to see now how the coefficient matrix of a bilinear form behaves under changes of coordinate systems. If $\mathbf{y} = C\mathbf{y}'$, and $\mathbf{x} = D\mathbf{x}'$, where C and D are invertible $n \times n$- and $m \times m$-matrices, we have: $\mathbf{x}^t = (\mathbf{x}')^t D^t$ and $\mathbf{x}^t B \mathbf{y} = (\mathbf{x}')^t D^t B C \mathbf{y}' = (\mathbf{x}')^t B' \mathbf{y}'$. That is, the coefficient matrix B is transformed into[7]

$$B' = D^t B C.$$

[7]Note that if B were the matrix of a linear map $\mathbb{K}^n \to \mathbb{K}^m$, the transformation rule would have been different: $B \mapsto D^{-1} B C$.

With a bilinear form $B : \mathcal{V} \times \mathcal{W} \to \mathbb{K}$, one can associate the **transposed** bilinear form $B^t : \mathcal{W} \times \mathcal{V} \to \mathbb{K}$ by simply changing the order of the arguments: $B^t(\mathbf{w}, \mathbf{v}) := B(\mathbf{v}, \mathbf{w})$. In coordinates, $B^t(\mathbf{f}_i, \mathbf{e}_j) = B(\mathbf{e}_j, \mathbf{f}_i)$, that is the coefficient matrix of B^t is transposed to that of B: $b^t_{ij} = b_{ji}$.

To a bilinear form $B : \mathcal{V} \times \mathcal{W} \to \mathbb{K}$, one can associate a linear map (denoted by the same letter) $B : \mathcal{W} \to \mathcal{V}^*$ from $\mathcal{W}$ to the dual of $\mathcal{V}$. Namely, to a vector $\mathbf{w} \in \mathcal{W}$, it associates the linear function on $\mathcal{V}$ which on each vector $\mathbf{v} \in \mathcal{V}$ assumes the value $B(\mathbf{v}, \mathbf{w})$. Conversely, given a linear map $B : \mathcal{W} \to \mathcal{V}^*$ it defines a bilinear form, whose value on the pair $(\mathbf{v}, \mathbf{w})$ is equal to the value of the linear function $B\mathbf{w} \in \mathcal{V}^*$ on the vector $\mathbf{v} \in \mathcal{V}$.

Choosing coordinate systems $\mathbf{v} = \sum x_i \mathbf{e}_i$ in $\mathcal{V}$, and $\mathbf{w} = \sum_j y_j \mathbf{f}_j$ in $\mathcal{W}$, one can describe the map $B : \mathcal{W} \to \mathcal{V}^*$ by its matrix with respect to the basis $\mathbf{f}_1, \dots, \mathbf{f}_n$ in $\mathcal{W}$, and $x_1, \dots, x_m$ in $\mathcal{V}^*$. We leave it as an exercise for the reader to check that $[b_{ij}]$ is the matrix (i.e. $b_{ij} = B(\mathbf{e}_i, \mathbf{f}_j)$). The map B^t, adjoint to $B : \mathcal{W} \to \mathcal{V}^*$, acts from $\mathcal{V}^{**} = \mathcal{V}$ to $\mathcal{W}^*$. Of course, it corresponds to the transposed bilinear form, and is described in coordinates by the transposed matrix.

In practice, we will be interested in the case when $\mathcal{V} = \mathcal{W}$, i.e. in bilinear forms of two vector arguments from the same space. A bilinear form $B : \mathcal{V} \times \mathcal{V} \to \mathbb{K}$ is called **symmetric** (**anti-symmetric**), if $B = B^t$ ($B = -B^t$), or more explicitly: $B(\mathbf{v}, \mathbf{w}) = B(\mathbf{w}, \mathbf{v})$ and $B(\mathbf{u}, \mathbf{v}) = -B(\mathbf{v}, \mathbf{u})$ for all vectors $\mathbf{u}, \mathbf{v}$ from $\mathcal{V}$. Every bilinear form on a vector space $\mathcal{V}$ can be uniquely written as the sum of a symmetric and anti-symmetric form:[8]

$$B(\mathbf{u}, \mathbf{v}) = \frac{B(\mathbf{u}, \mathbf{v}) + B(\mathbf{v}, \mathbf{u})}{2} + \frac{B(\mathbf{u}, \mathbf{v}) - B(\mathbf{v}, \mathbf{u})}{2}.$$

Using a basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ in $\mathcal{V}$ (and the same basis $\mathbf{f}_j = \mathbf{e}_j$ in $\mathcal{W} = \mathcal{V}$), one obtains the coefficient matrix $[b_{ij}]$ of the bilinear form, which is a square matrix with the entries $b_{ij} = B(\mathbf{e}_i, \mathbf{e}_j)$. The matrix of a bilinear form is **symmetric**, i.e. $b_{ij} = b_{ij}$, or **anti-symmetric**, i.e. $b_{ij} = -b_{ji}$, whenever the bilinear form is.

Under a change of coordinates, $\mathbf{x} = C\mathbf{x}'$, $\mathbf{y} = C\mathbf{y}'$, the coefficient matrix B is transformed into

$$B' = C^t B C,$$

and remains symmetric (anti-symmetric), whenever B is.

[8] Note that division by 2 is used here, which will not work over the field containing $\mathbb{Z}_2 = \{0, 1\}$, where $1 + 1 = 0$.

EXERCISES

186. Find the coefficient matrix of the dot product. ✓

187. Prove that all anti-symmetric bilinear forms in $\mathbb{R}^2$ are proportional to each other. ϟ

188. Represent the bilinear form $B = 2x_1(y_1 + y_2)$ in $\mathbb{R}^2$ as the sum $S + A$ of symmetric and anti-symmetric ones. ✓

189. Is AB necessarily symmetric if A and B are? ✓

190. Prove that for any matrix A, both A^tA and AA^t are symmetric.

191. Find a square matrix A such that $A^tA \neq AA^t$. ϟ

Quadratic Forms

From a bilinear form $B : \mathcal{V} \times \mathcal{V} \to \mathbb{K}$, one obtains a **quadratic form** $Q_B : \mathcal{V} \to \mathbb{K}$ by "restricting to the diagonal" $\mathcal{V} \subset \mathcal{V} \times \mathcal{V}$:

$$Q_B(\mathbf{v}) := B(\mathbf{v}, \mathbf{v}).$$

If B is written as the sum $B = S+A$ of symmetric and anti-symmetric bilinear forms, we find that $A(\mathbf{v}, \mathbf{v}) = -A(\mathbf{v}, \mathbf{v}) = 0$, and so $Q_B(\mathbf{v}) = Q_S(\mathbf{v}) = S(\mathbf{v}, \mathbf{v})$.

A symmetric bilinear form can be reconstructed from the corresponding quadratic form:

$$S(\mathbf{u}, \mathbf{v}) = \frac{S(\mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v}) - S(\mathbf{u}, \mathbf{u}) - S(\mathbf{v}, \mathbf{v})}{2}.$$

In fact, one can define a quadratic form as a function $Q : \mathcal{V} \to \mathbb{K}$ which is *homogeneous of degree* 2 (i.e. $Q(\lambda\mathbf{v}) = \lambda^2 Q(\mathbf{v})$), and such that $S(\mathbf{u}, \mathbf{v}) := (Q(\mathbf{u} + \mathbf{v}) - Q(\mathbf{u}) - Q(\mathbf{v}))/2$ is bilinear.

In coordinates, any $n \times n$-matrix $Q = [q_{ij}]$, which is symmetric, defines a symmetric bilinear form $\mathbf{x}^tQ\mathbf{y}$ and the corresponding quadratic form

$$Q(\mathbf{x}) = \mathbf{x}^tQ\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i q_{ij} x_j.$$

EXERCISES

192. Is the function xy: linear? bilinear? quadratic? ✓

193. Find the symmetric bilinear forms corresponding to the quadratic forms $(x_1 + \cdots + x_n)^2$ and $\sum_{i<j} x_i x_j$. ✓

Hermitian Forms

Here we need to assume that $\mathbb{K} = \mathbb{C}$, the field of complex numbers, since the operation $\lambda \mapsto \bar{\lambda}$ of complex conjugation of scalars will be involved.

A **sesquilinear form** on a complex vector space $\mathcal{V}$ is defined as a function $T : \mathcal{V} \times \mathcal{V} \to \mathbb{C}$ of two vector arguments which is **anti-linear** (or **half-linear**) in the first argument and linear in the second. By definition, this means that for all $\lambda, \mu \in \mathbb{C}$ and $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{V}$

$$T(\lambda\mathbf{u} + \mu\mathbf{v}, \mathbf{w}) = \bar{\lambda}T(\mathbf{u}, \mathbf{w}) + \bar{\mu}T(\mathbf{v}, \mathbf{w}),$$
$$T(\mathbf{w}, \lambda\mathbf{u} + \mu\mathbf{v}) = \lambda T(\mathbf{w}, \mathbf{u}) + \mu T(\mathbf{w}, \mathbf{v}).$$

In coordinates, let $\mathbf{v} = z_1\mathbf{e}_1 + \cdots + z_n\mathbf{e}_n$ and $\mathbf{w} = w_1\mathbf{e}_1 + \cdots + w_n\mathbf{e}_n$, where $z_i, w_j \in \mathbb{C}$. Then

$$T(\mathbf{v}, \mathbf{w}) = \sum_{i=1}^{n}\sum_{j=1}^{n} \overline{z_i}\, t_{ij} w_j, \quad \text{where} \quad t_{ij} = T(\mathbf{e}_i, \mathbf{e}_j).$$

Conversely, given an arbitrary complex $n \times n$-matrix $[t_{ij}]$, the above formula defines a sesquilinear form on the coordinate space $\mathcal{V} = \mathbb{C}^n$.

Given a sesquilinear form T, its **Hermitian adjoint**[9] (or simply **adjoint** form $T^\dagger$ is defined by

$$T^\dagger(\mathbf{u}, \mathbf{v}) := \overline{T(\mathbf{v}, \mathbf{u})} \quad \text{for all } \mathbf{u}, \mathbf{v} \in \mathcal{V}.$$

Note that it is also sesquilinear, that is, anti-linear with respect to $\mathbf{u}$ and linear with respect to $\mathbf{v}$. In coordinates, the coefficient matrix $[t^\dagger_{ij}]$ of the adjoint form and the matrix $[t_{ij}]$ are **Hermitian conjugate**, i.e., are obtained from each other by the operations of complex conjugation of all entries and transposition. Indeed,

$$t^\dagger_{ij} = T^\dagger(\mathbf{e}_i, \mathbf{e}_j) = \overline{T(\mathbf{e}_j, \mathbf{e}_i)} = \overline{t_{ji}}.$$

A sesquilinear form T is called **Hermitian-symmetric** if $T^\dagger = T$ and **Hermitian-anti-symmetric** if $T^\dagger = -T$. In fact, T is Hermitian symmetric whenever $\sqrt{-1}T$ is Hermitian anti-symmetric (check this). Moreover, every sesquilinear form can be uniquely written as the sum of Hermitian-symmetric and Hermitian-anti-symmetric ones:

$$T = \frac{T + T^\dagger}{2} + i\frac{T - T^\dagger}{2i}, \quad i = \sqrt{-1}.$$

[9] After French mathematician Charles **Hermite** (1822–1901).

Similarly to the case of quadratic forms, a sesquilinear form T can be reconstructed from its restriction $\mathbf{v} \mapsto T(\mathbf{v}, \mathbf{v})$ to the diagonal (see Exercises). This restriction is a complex-valued function of one vector variable. It is real-homogeneous of degree 2 in the following sense: $T(\lambda\mathbf{v}, \lambda\mathbf{v}) = |\lambda|^2 T(\mathbf{v}, \mathbf{v})$. When T is Hermitian-symmetric, this function takes on purely real values, and when T is Hermitian-anti-symmetric, purely imaginary. We will call such functions **Hermitian** and **anti-Hermitian quadratic forms**, or simply an **Hermitian** and **anti-Hermitian forms**.

In coordinates, an Hermitian quadratic form looks this way:

$$H(\mathbf{z}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \overline{z_i}\, h_{ij} z_j, \quad \text{where} \quad h_{ji} = \overline{h_{ij}}.$$

When $[h_{ij}]$ is the identity matrix, the Hermitian form is $|z_1|^2 + \cdots + |z_n|^n$. The space $\mathbb{C}^n$ equipped with this Hermitian form is called the **standard Hermitian space**, and plays in complex geometry the role analogous to the **standard Euclidean space** of real geometry.

EXERCISES

194. Which of the function $zw, \bar{z}w, z\bar{w}, \bar{z}\bar{w}$ of two complex variables are: (a) linear with respect to z? (b) anti-linear with respect to w?, (c) bilinear? (d) sesquilinear?

195. Let $T(z_1, z_2; w_1, w_2) = \overline{z_1}\, w_2$. Compute $T^\dagger$, represent T as the sum $S_1 + iS_2$, where S_1, S_2 are Hermitian symmetric, and compute the Hermitian quadratic forms H_1 and H_2 corresponding to S_1 and S_2. ✓

196. Verify that $\mathrm{Re}(\overline{z_1}\, z_2 + \cdots + \overline{z_{n-1}}\, z_n)$ is an Hermitian quadratic form and find the corresponding Hermitian-symmetric sesquilinear form. ✓

197. Determine how the matrix of a sesquilinear form $T : \mathcal{V} \times \mathcal{W} \to \mathbb{C}$ is transformed under the coordinate changes $\mathbf{v} = C\mathbf{v}'$ and $\mathbf{w} = D\mathbf{w}'$. ✓

198. Show that all sesquilinear forms in $\mathbb{C}^n$ form a complex vector space, and find its dimension.

199. Do the Hermitian symmetric and Hermitian-anti-symmetric sesquilinear forms in $\mathbb{C}^n$ form a complex vector space? a real vector space? Find their dimensions. ✓

200. Let S be an Hermitian-symmetric sesquilinear form. Show that

$$S(\mathbf{z} + \mathbf{w}, \mathbf{z} + \mathbf{w}) = S(\mathbf{z}, \mathbf{z}) + S(\mathbf{w}, \mathbf{w}) + 2\,\mathrm{Re}\, S(\mathbf{z}, \mathbf{w})$$
$$S(\mathbf{z} + i\mathbf{w}, \mathbf{z} + i\mathbf{w}) = S(\mathbf{z}, \mathbf{z}) + S(\mathbf{w}, \mathbf{w}) - 2\,\mathrm{Im}\, S(\mathbf{z}, \mathbf{w})$$

and derive from this, that S is uniquely determined by the corresponding Hermitian quadratic form $H(\mathbf{z}) := S(\mathbf{z}, \mathbf{z})$.

3 Determinants

Definition

Let A be a *square* matrix of size n:

$$A = \begin{bmatrix} a_{11} & ... & a_{1n} \\ & ... & \\ a_{n1} & ... & a_{nn} \end{bmatrix}$$

Its **determinant** is a *scalar* $\det A$ defined by the formula

$$\det A = \sum_{\sigma} \varepsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} ... a_{n\sigma(n)}.$$

Here σ is a **permutation** of the indices $1, 2, ..., n$. A permutation σ can be considered as an invertible function $i \mapsto \sigma(i)$ from the set of n elements $\{1, ..., n\}$ to itself. We use the functional notation $\sigma(i)$ in order to specify the i-th term in the permutation $\sigma = \begin{pmatrix} 1 & \dots & n \\ \sigma(1) & \dots & \sigma(n) \end{pmatrix}$. Thus, each **elementary product** in the determinant formula contains exactly one matrix entry from each row, and these entries are chosen from n different columns. The sum is taken over all $n!$ ways of making such choices. The coefficient $\varepsilon(\sigma)$ in front of the elementary product equals 1 or -1 and is called the **sign** of the permutation σ.

We will explain the general rule of the signs after a few examples. In these examples, we begin using one more conventional notation for determinants. According to it, a square array of matrix entries placed between two vertical bars denotes the *determinant* of the matrix. Thus, $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ denotes a *matrix*, but $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$ denotes a *number* equal to the determinant of that matrix.

`Examples.` (1) For $n = 1$, the determinant $|a_{11}| = a_{11}$.

(2) For $n = 2$, we have: $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$.

(3) For $n = 3$, we have $3! = 6$ summands

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} =$$

$a_{11}a_{22}a_{33} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} - a_{13}a_{22}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32}$ corresponding to permutations $\binom{123}{123}$, $\binom{123}{213}$, $\binom{123}{231}$, $\binom{123}{321}$, $\binom{123}{312}$, $\binom{123}{132}$.

The rule of signs for $n = 3$ is schematically shown on Figure 27.

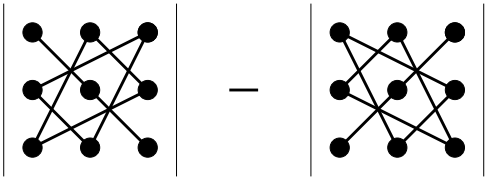

Figure 28

EXERCISES

201. Prove that the following determinant is equal to 0:

$$\begin{vmatrix} 0 & 0 & 0 & a & b \\ 0 & 0 & 0 & c & d \\ 0 & 0 & 0 & e & f \\ p & q & r & s & t \\ v & w & x & y & z \end{vmatrix}. \qquad \text{ϟ}$$

202. Compute determinants:

$$\begin{vmatrix} \mathbf{cos}\, x & -\mathbf{sin}\, x \\ \mathbf{sin}\, x & \mathbf{cos}\, x \end{vmatrix}, \begin{vmatrix} \mathbf{cosh}\, x & \mathbf{sinh}\, x \\ \mathbf{sinh}\, x & \mathbf{cosh}\, x \end{vmatrix}, \begin{vmatrix} \mathbf{cos}\, x & \mathbf{sin}\, y \\ \mathbf{sin}\, x & \mathbf{cos}\, y \end{vmatrix}. \qquad \checkmark$$

203. Compute determinants:

$$\begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{vmatrix}, \begin{vmatrix} 0 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 6 \end{vmatrix}, \begin{vmatrix} 1 & i & 1+i \\ -i & 1 & 0 \\ 1-i & 0 & 1 \end{vmatrix}. \qquad \checkmark$$

Parity of Permutations

The general rule of signs relies on properties of permutations.

Let Δ_n denote the following polynomial in n variables $x_1, \ldots, x_n$:

$$\Delta_n(x_1, \ldots, x_n) = \prod_{1 \le i < j \le n} (x_i - x_j).$$

`Examples:` $\Delta_2 = x_1 - x_2$, $\Delta_3 = (x_1 - x_2)(x_1 - x_3)(x_2 - x_3)$. By definition, $\Delta_1 = 1$. In general, Δ_n is the product of all "n-choose-2" linear factors $x_i - x_j$ written in such a way that $i < j$.

Let σ be a permutation of $\{1,\dots,n\}$. It acts on polynomials P in the variables $x_1,\dots,x_n$ by permutation of the variables: $(\sigma P)(x_1,\dots,x_n) := P(x_{\sigma(1)},\dots,x_{\sigma(n)})$.

`Example.` Let $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$. Then

$$\sigma\Delta_3 = (x_3-x_1)(x_3-x_2)(x_1-x_2) = (-1)^2(x_1-x_3)(x_2-x_3)(x_1-x_2).$$

One says that σ **inverses** a pair of indices $i<j$ if $\sigma(i)>\sigma(j)$. The total number $l(\sigma)$ of pairs $i<j$ that σ inverses is called the **length** of the permutation σ. Thus, in the previous example, σ inverses the pairs $(1,2)$ and $(1,3)$, and has length $l(\sigma)=2$.

`Lemma.` $\sigma\Delta_n = \varepsilon(\sigma)\Delta_n$, ***where*** $\varepsilon(\sigma) = (-1)^{l(\sigma)}$.

`Proof.` Indeed, a permutation of $\{1,\dots,n\}$ also permutes all pairs $i\neq j$, and hence permutes all the linear factors in Δ_n. However, a factor x_i-x_j is transformed into $x_{\sigma(i)}-x_{\sigma(j)}$, which occurs in the product Δ_n with the same sign whenever $\sigma(i)<\sigma(j)$, and with the opposite sign whenever $\sigma(i)>\sigma(j)$. Thus, $\sigma\Delta_n$ differs from Δ_n by the sign $(-1)^{l(\sigma)}$. □

A permutation σ is called **even** or **odd** depending on the sign $\varepsilon(\sigma)$, i.e. when the length is even or odd respectively.

`Examples.` (1) The **identity permutation** id (defined by $\mathrm{id}(i)=i$ for all i) is even since $l(\mathrm{id})=0$.

(2) Consider a **transposition** τ, i.e. a permutation that swaps two indices, say $i<j$, leaving all other indices in their respective places. Then $\tau(j)<\tau(i)$, i.e. τ inverses the pair of indices $i<j$. Besides, for every index k such that $i<k<j$ we have: $\tau(j)<\tau(k)<\tau(i)$, i.e. both pairs $i<k$ and $k<j$ are inverted. Note that all other pairs of indices are not inverted by τ, and hence $l(\tau)=2(j-i)+1$. In particular, *every transposition is odd*: $\varepsilon(\tau)=-1$.

`Proposition.` ***Composition of two even or two odd permutations is even, and composition of one even and one odd permutation is odd:***

$$\varepsilon(\sigma\sigma') = \varepsilon(\sigma)\varepsilon(\sigma').$$

`Proof.` We have:

$$\varepsilon(\sigma\sigma')\Delta_n := (\sigma\sigma')\Delta_n = \sigma(\sigma'\Delta_n) = \varepsilon(\sigma')\sigma\Delta_n = \varepsilon(\sigma')\varepsilon(\sigma)\Delta_n.$$

`Corollary 1.` ***Inverse permutations have the same parity.***

Corollary 2. ***Whenever a permutation is written as the product of transpositions, the parity of the number of the transpositions in the product remains the same and coincides with the parity of the permutation: If*** $\sigma = \tau_1 \dots \tau_N$, ***then*** $\varepsilon(\sigma) = (-1)^N$.

Here are some illustrations of the above properties in connection with the definition of determinants.

Examples. (3) The transposition (21) is odd. That is why the term $a_{12}a_{21}$ occurs in 2×2-determinants with the negative sign.

(4) The permutations $\binom{123}{123}$, $\binom{123}{213}$, $\binom{123}{231}$, $\binom{123}{321}$, $\binom{123}{312}$, $\binom{123}{132}$ have lengths $l = 0, 1, 2, 3, 2, 1$ and respectively signs $\varepsilon = 1, -1, 1, -1, 1, -1$ (thus explaining Figure 27). Notice that each next permutation here is obtained from the previous one by an extra flip.

(5) The permutation $\binom{1234}{4321}$ inverses all the 6 pairs of indices and has therefore length $l = 6$. Thus the elementary product $a_{14}a_{23}a_{32}a_{41}$ occurs with the sign $\varepsilon = (-1)^6 = +1$ in the definition of 4×4-determinants.

(6) Since inverse permutations have the same parity, the definition of determinants can be rewritten "by columns:"

$$\det A = \sum_{\sigma} \varepsilon(\sigma) a_{\sigma(1)1} \dots a_{\sigma(n)n}.$$

Indeed, each summand in this formula is equal to the summand in the original definition corresponding to the permutation σ^{-1}, and *vice versa*. Namely, reordering the factors $a_{\sigma(1)1} \dots a_{\sigma(n)n}$, so that $\sigma(1), \dots, \sigma(n)$ increase monotonically, yields $a_{1\sigma^{-1}(1)} \dots a_{n\sigma^{-1}(n)}$.

EXERCISES

204. List all the 24 permutations of $\{1, 2, 3, 4\}$, find the length and the sign of each of them. ϟ

205. Find the length of the following permutation:

$$\begin{pmatrix} 1 & 2 & \dots & k & k+1 & k+2 & \dots & 2k \\ 1 & 3 & \dots & 2k-1 & 2 & 4 & \dots & 2k \end{pmatrix}. \qquad \checkmark$$

206. Find the maximal possible length of permutations of $\{1, \dots, n\}$. ϟ

207. Find the length of a permutation $\begin{pmatrix} 1 & \dots & n \\ i_1 & \dots & i_n \end{pmatrix}$ given the length l of the permutation $\begin{pmatrix} 1 & \dots & n \\ i_n & \dots & i_1 \end{pmatrix}$. ✓

208. Prove that inverse permutations have the same length. ϟ

209. Compare parities of permutations of the letters *a,g,h,i,l,m,o,r,t* in the words *logarithm* and *algorithm*. ϟ

210. Prove that the identity permutations are the only ones of length 0.

211. Find all permutations of length 1. ✓

212.* Show that every permutation σ can be written as the product of $l(\sigma)$ transpositions of nearby indices. ϟ

213.* Represent the permutation $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 5 & 1 & 3 & 2 \end{pmatrix}$ as composition of a minimal number of transpositions. ✓

214. Do products $a_{13}a_{24}a_{53}a_{41}a_{35}$ and $a_{21}a_{13}a_{34}a_{55}a_{42}$ occur in the defining formula for determinants of size 5? ✓

215. Find the signs of the elementary products $a_{23}a_{31}a_{42}a_{56}a_{14}a_{65}$ and $a_{32}a_{43}a_{14}a_{51}a_{66}a_{25}$ in the definition of determinants of size 6 by computing the numbers of inverted pairs of indices. ✓

Properties of determinants

(i) ***Transposed matrices have equal determinants:***

$$\det A^t = \det A.$$

This follows from the last Example. Below, we will think of an $n \times n$ matrix as an array $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$ of its n columns of size n (vectors from $\mathbb{C}^n$ if you wish) and formulate all further properties of determinants in terms of columns. The same properties hold true for rows, since the transposition of A changes columns into rows without changing the determinant.

(ii) ***Interchanging any two columns changes the sign of the determinant:***

$$\det[..., \mathbf{a}_j, ..., \mathbf{a}_i, ...] = -\det[..., \mathbf{a}_i, ..., \mathbf{a}_j, ...].$$

Indeed, the operation replaces each permutation in the definition of determinants by its composition with the transposition of the indices i and j. Thus changes the parity of the permutation, and thus reverses the sign of each summand.

Rephrasing this property, one says that the determinant, considered as a function of n vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$ is **totally anti-symmetric**, i.e. changes the sign under every odd permutation of the vectors, and stays invariant under even. It implies that *a matrix with two equal columns has zero determinant.* It also allows one to formulate further

column properties of determinants referring to the 1st column only, since the properties of all columns are alike.

(iii) ***Multiplication of a column by a number multiplies the determinant by this number:***

$$\det[\lambda \mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n] = \lambda \det[\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n].$$

Indeed, this operation simply multiplies each of the $n!$ elementary products by the factor of λ.

This property shows that *a matrix with a zero column has zero determinant.*

(iv) ***The determinant function is*** **additive** ***with respect to each column:***

$$\det[\mathbf{a}_1' + \mathbf{a}_1'', \mathbf{a}_2, ..., \mathbf{a}_n] = \det[\mathbf{a}_1', \mathbf{a}_2, ..., \mathbf{a}_n] + \det[\mathbf{a}_1'', \mathbf{a}_2, ..., \mathbf{a}_n].$$

Indeed, each elementary product contains exactly one factor picked from the 1-st column and thus splits into the sum of two elementary products $a'_{\sigma(1)1}a_{\sigma(2)2}...a_{\sigma(n)n}$ and $a''_{\sigma(1)1}a_{\sigma(2)2}...a_{\sigma(n)n}$. Summing up over all permutations yields the sum of two determinants on the right hand side of the formula.

The properties (*iv*) and (*iii*) together mean that *the determinant function is linear with respect to each column* separately. Together with the property (*ii*), they show that ***adding a multiple of one column to another one does not change the determinant of the matrix***. Indeed,

$$|\mathbf{a}_1 + \lambda \mathbf{a}_2, \mathbf{a}_2, ...| = |\mathbf{a}_1, \mathbf{a}_2, ...| + \lambda\, |\mathbf{a}_2, \mathbf{a}_2, ...| = |\mathbf{a}_1, \mathbf{a}_2, ...|\,,$$

since the second summand has two equal columns.

The determinant function shares all the above properties with the identically zero function. The following property shows that these functions do not coincide.

(v) $\det I = 1$.

Indeed, since all off-diagonal entries of the identity matrix are zeroes, the only elementary product in the definition of $\det A$ that survives is $a_{11}...a_{nn} = 1$.

The same argument shows that *the determinant of any diagonal matrix equals the product of the diagonal entries.* It is not hard to generalize the argument in order to see that the determinant of any

upper or lower triangular matrix is equal to the product of the diagonal entries. One can also deduce this from the following factorization property valid for block triangular matrices.

Consider an $n \times n$-matrix $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ subdivided into four **blocks** A, B, C, D of sizes $m \times m,\ m \times l,\ l \times m$ and $l \times l$ respectively (where of course $m + l = n$). We will call such a matrix **block triangular** if C or B is the zero matrix 0. We claim that

$$\det \begin{bmatrix} A & B \\ 0 & D \end{bmatrix} = \det A \ \det D.$$

Indeed, consider a permutation σ of $\{1, ..., n\}$ which sends at least one of the indices $\{1, ..., m\}$ to the other part of the set, $\{m+1, ..., m+l\}$. Then σ must send at least one of $\{m+1, ..., m+l\}$ back to $\{1, ..., m\}$. This means that every elementary product in our $n \times n$-determinant which contains a factor from B must also contain a factor from C, and hence vanish, if $C = 0$. Thus only the permutations σ which permute $\{1, ..., m\}$ separately from $\{m+1, ..., m+l\}$ contribute to the determinant in question. Elementary products corresponding to such permutations factor into elementary products from $\det A$ and $\det D$ and eventually add up to the product $\det A \det D$.

Of course, the same holds true if $B = 0$ instead of $C = 0$.

We will use the factorization formula in the 1st proof of the following fundamental property of determinants.

EXERCISES

216. Compute the determinants

$$\begin{vmatrix} 13247 & 13347 \\ 28469 & 28569 \end{vmatrix}, \quad \begin{vmatrix} 246 & 427 & 327 \\ 1014 & 543 & 443 \\ -342 & 721 & 621 \end{vmatrix}. \qquad \checkmark$$

217. The numbers 195, 247, and 403 are divisible by 13. Prove that the following determinant is also divisible by 13: $\begin{vmatrix} 1 & 9 & 5 \\ 2 & 4 & 7 \\ 4 & 0 & 3 \end{vmatrix}$. ϟ

218. Professor Dumbel writes his office and home phone numbers as a 7×1-matrix O and 1×7-matrix H respectively. Help him compute $\det(OH)$. ✓

219. How does a determinant change if all its n columns are rewritten in the opposite order? ✓

220.* Solve the equation $\begin{vmatrix} 1 & x & x^2 & ... & x^n \\ 1 & a_1 & a_1^2 & ... & a_1^n \\ 1 & a_2 & a_2^2 & ... & a_2^n \\ & & & ... & \\ 1 & a_n & a_n^2 & ... & a_n^n \end{vmatrix} = 0$, where all $a_1, ..., a_n$ are given distinct numbers. ✓

221. Prove that an anti-symmetric matrix of size n has zero determinant if n is odd. ϟ

Multiplicativity

`Theorem.` ***The determinant is multiplicative with respect to matrix products: for arbitrary*** $n \times n$***-matrices*** A ***and*** B***,***

$$\det(AB) = (\det A)(\det B).$$

We give two proofs: one *ad hoc*, the other more conceptual.

`Proof I.` Consider the auxiliary $2n \times 2n$ matrix $\begin{bmatrix} A & 0 \\ -I & B \end{bmatrix}$ with the determinant equal to the product $(\det A)(\det B)$ according to the factorization formula. We begin to change the matrix by adding to the last n columns linear combinations of the first n columns with such coefficients that the submatrix B is eventually replaced by zero submatrix. Thus, in order to kill the entry b_{kj} we must add the b_{kj}-multiple of the k-th column to the $n + j$-th column. According to the properties of determinants (see (iv)) these operations do not change the determinant but transform the matrix to the form $\begin{bmatrix} A & C \\ -I & 0 \end{bmatrix}$. We ask the reader to check that the entry c_{ij} of the submatrix C in the upper right corner equals $a_{i1}b_{1j} + ... + a_{in}b_{nj}$ so that $C = AB$ is the matrix product! Now, interchanging the i-th and $n + i$-th columns, $i = 1, ..., n$, we change the determinant by the factor of $(-1)^n$ and transform the matrix to the form $\begin{bmatrix} C & A \\ 0 & -I \end{bmatrix}$. The factorization formula applies again and yields $\det C \det(-I)$. We conclude that $\det C = \det A \det B$ since $\det(-I) = (-1)^n$ compensates for the previous factor $(-1)^n$.□

`Proof II.` We will first show that the properties (i – v) completely characterize $\det[\mathbf{v}_1, \ldots, \mathbf{v}_n]$ as a function of n columns $\mathbf{v}_i$ of size n.

Indeed, consider a function f, which to n columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$, associates a number $f(\mathbf{v}_1, \ldots, \mathbf{v}_n)$. Suppose that f is *linear* with

respect to each column. Let $\mathbf{e}_i$ denote the ith column of the identity matrix. Since $\mathbf{v}_1 = \sum_{i=1}^n v_{i1}\mathbf{e}_i$, we have:

$$f(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = \sum_{i=1}^n v_{i1} f(\mathbf{e}_i, \mathbf{v}_2, \ldots, \mathbf{v}_n).$$

Using linearity with respect to the 2nd column $\mathbf{v}_2 = \sum_{j=1}^n v_{j2}\mathbf{e}_j$, we similarly obtain:

$$f(\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n) = \sum_{i=1}^n \sum_{j=1}^n v_{i1} v_{j2} f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{v}_3, \ldots, \mathbf{v}_n).$$

Proceeding the same way with all columns, we get:

$$f(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \sum_{i_1, \ldots, i_n} v_{i_1 1} \cdots v_{i_n n} f(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_n}).$$

Thus, f is determined by its values $f(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_n})$ on strings of n basis vectors.

Let us assume now that f is *totally anti-symmetric*. Then, if any two of the indices $i_1, \ldots, i_n$ coincide, we have: $f(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_n}) = 0$. All other coefficients correspond to *permutations* $\sigma = \begin{pmatrix} 1 & \ldots & n \\ i_1 & \ldots & i_n \end{pmatrix}$ of the indices $(1, \ldots, n)$, and hence satisfy:

$$f(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_n}) = \varepsilon(\sigma) f(\mathbf{e}_1, \ldots, \mathbf{e}_n).$$

Therefore, we find:

$$\begin{aligned} f(\mathbf{v}_1, \ldots, \mathbf{v}_n) &= \sum_\sigma v_{\sigma(1)1} \ldots v_{\sigma(n)n} \varepsilon(\sigma) f(\mathbf{e}_1, \ldots, \mathbf{e}_n), \\ &= f(\mathbf{e}_1, \ldots, \mathbf{e}_n) \det[\mathbf{v}_1, \ldots, \mathbf{v}_n]. \end{aligned}$$

Thus, we have established:

`Proposition 1.` ***Every totally anti-symmetric function of n coordinate vectors of size n which is linear in each of them is proportional to the determinant function.***

Next, given an $n \times n$ matrix C, put

$$f(\mathbf{v}_1, \ldots, \mathbf{v}_n) := \det[C\mathbf{v}_1, \ldots, C\mathbf{v}_n].$$

Obviously, the function f is totally anti-symmetric in all $\mathbf{v}_i$ (since det is). Multiplication by C is linear:

$$C(\lambda\mathbf{u} + \mu\mathbf{v}) = \lambda C\mathbf{u} + \mu C\mathbf{v} \quad \text{for all } \mathbf{u}, \mathbf{v} \text{ and } \lambda, \mu.$$

Therefore, f is linear with respect to each $\mathbf{v}_i$ (as composition of two linear operations). By the previous result, f is proportional to det. Since $C\mathbf{e}_i$ are columns of C, we conclude that the coefficient of proportionality $f(\mathbf{e}_1, \dots, \mathbf{e}_n) = \det C$. Thus, we have found the following interpretation of $\det C$.

`Proposition 2.` $\det C$ ***is the factor by which the determinant function of*** n ***vectors*** $\mathbf{v}_i$ ***is multiplied when the vectors are replaced with*** $C\mathbf{v}_i$***.***

Now our theorem follows from the fact that when $C = AB$, the substitution $\mathbf{v} \mapsto C\mathbf{v}$ is the composition $\mathbf{v} \mapsto A\mathbf{v} \mapsto AB\mathbf{v}$ of consecutive substitutions defined by A and B. Under the action of A, the function det is multiplied by the factor $\det A$, then under the action of B by another factor $\det B$. But the resulting factor $(\det A)(\det B)$ must be equal to $\det C$. □

`Corollary.` ***If*** A ***is invertible, then*** $\det A$ ***is invertible.***

Indeed, $(\det A)(\det A^{-1}) = \det I = 1$, and hence $\det A^{-1}$ is reciprocal to $\det A$. The converse statement: that matrices with invertible determinants are invertible, is also true due to the explicit formula for the inverse matrix, described in the next section.

`Remark.` Of course, a real or complex number $\det A$ is invertible whenever $\det A \neq 0$. Yet over the integers $\mathbb{Z}$ this is not the case: the only invertible integers are ± 1. The above formulation, and several similar formulations that follow, which refer to invertibility of determinants, are preferable as they are more general.

EXERCISES

222. How do similarity transformations of a given matrix affect its determinant? ✓

223. Prove that the sign of the determinant of the coefficient marix of a real quadratic form does not depend on the coordinate system. ϟ

The Cofactor Theorem

In the determinant formula for an $n \times n$-matrix A each elementary product $\pm a_{1\sigma(1)} \dots$ begins with one of the entries $a_{11}, \dots, a_{1n}$ of the first row. The sum of all terms containing a_{11} in the 1-st place is the product of a_{11} with the determinant of the $(n-1) \times (n-1)$-matrix obtained from A by crossing out the 1-st row and the 1-st column. Similarly, the sum of all terms containing a_{12} in the 1-st place looks like the product of a_{12} with the determinant obtained by

crossing out the 1-st row and the 2-nd column of A. In fact it differs by the factor of -1 from this product, since switching the columns 1 and 2 changes signs of all terms in the determinant formula and interchanges the roles of a_{11} and a_{12}. Proceeding in this way with $a_{13}, ..., a_{1n}$ we arrive at the **cofactor expansion** formula for $\det A$ which can be stated as follows.

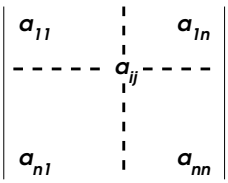

Figure 29

i \ j	1	2	3	4	5
1	+	−	+	−	+
2	−	+	−	+	−
3	+	−	+	−	+
4	−	+	−	+	−
5	+	−	+	−	+

Figure 30

The determinant of the $(n-1) \times (n-1)$-matrix obtained from A by crossing out the row i and column j is called the (ij)**-minor** of A (Figure 28). Denote it by M_{ij}. The (ij)**-cofactor** A_{ij} of the matrix A is the number that differs from the minor M_{ij} by a factor ± 1:

$$A_{ij} = (-1)^{i+j} M_{ij}.$$

The chess-board of the signs $(-1)^{i+j}$ is shown on Figure 29. With these notations, the cofactor expansion formula reads:

$$\det A = a_{11}A_{11} + a_{12}A_{12} + ... + a_{1n}A_{1n}.$$

Example.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

Using the properties (i) and (ii) of determinants we can adjust the cofactor expansion to the i-th row or j-th column:

$$\det A = a_{i1}A_{i1} + ... + a_{in}A_{in} = a_{1j}A_{1j} + ... + a_{nj}A_{nj}, \ i, j = 1, ..., n.$$

These formulas reduce evaluation of $n \times n$-determinants to that of $(n-1) \times (n-1)$-determinants and can be useful in recursive computations.

Furthermore, we claim that applying the cofactor formula to the entries of the i-th row but picking the cofactors of another row we get the zero sum:

$$a_{i1}A_{j1} + ... + a_{in}A_{jn} = 0 \text{ if } i \neq j.$$

Indeed, construct a new matrix $\tilde{A}$ replacing the j-th row by a copy of the i-th row. This forgery does not change the cofactors $A_{j1}, ..., A_{jn}$ (since the j-th row is crossed out anyway) and yields the cofactor expansion $a_{i1}A_{j1} + ... + a_{in}A_{jn}$ for $\det \tilde{A}$. But $\tilde{A}$ has two identical rows and hence $\det \tilde{A} = 0$. The same arguments applied to the columns yield the dual statement:

$$a_{1i}A_{1j} + ... + a_{ni}A_{nj} = 0 \text{ if } i \neq j.$$

All the above formulas can be summarized in a single matrix identity. Introduce the $n \times n$-matrix $\operatorname{adj}(A)$, called **adjugate** to A, by placing the cofactor A_{ij} on the intersection of j-th row and i-th column. In other words, each a_{ij} is replaced with the corresponding cofactor A_{ij}, and then the resulting matrix is transposed:

$$\operatorname{adj} \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \dots & a_{ij} & \dots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} = \begin{bmatrix} A_{11} & \dots & A_{n1} \\ \dots & A_{ji} & \dots \\ A_{1n} & \dots & A_{nn} \end{bmatrix}.$$

`Theorem.` $A \ \operatorname{adj}(A) = (\det A)\ I = \operatorname{adj}(A)\ A$.

`Corollary.` ***If*** $\det A$ ***is invertible then*** A ***is invertible, and***

$$A^{-1} = \frac{1}{\det A} \operatorname{adj}(A).$$

`Example.` If $ad - bc \neq 0$, then $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

EXERCISES

224. Prove that the adjugate matrix of an upper (lower) triangular matrix is upper (lower) triangular.

225. Which triangular matrices are invertible?

226. Compute the determinants: ($*$ is a wild card):

$$(a) \begin{vmatrix} * & * & * & a_n \\ * & * & \dots & 0 \\ * & a_2 & 0 & \dots \\ a_1 & 0 & \dots & 0 \end{vmatrix}, \quad (b) \begin{vmatrix} * & * & a & b \\ * & * & c & d \\ e & f & 0 & 0 \\ g & h & 0 & 0 \end{vmatrix}. \qquad \checkmark$$

227. Compute determinants using cofactor expansions:

$$(a)\ \begin{vmatrix} 1 & 2 & 2 & 1 \\ 0 & 1 & 0 & 2 \\ 2 & 0 & 1 & 1 \\ 0 & 2 & 0 & 1 \end{vmatrix}, \quad (b)\ \begin{vmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{vmatrix}. \qquad \checkmark$$

228. Compute inverses of matrices using the Cofactor Theorem:

$$(a)\ \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 2 & 3 & 1 \end{bmatrix}, \quad (b)\ \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}. \qquad \checkmark$$

229. Solve the systems of linear equations $A\mathbf{x} = \mathbf{b}$ where A is one of the matrices of the previous exercise, and $\mathbf{b} = [1, 0, 1]^t$. ✓

230. Compute

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

231. Express $\det(\mathrm{adj}(A))$ of the adjugate matrix via $\det A$. ✓

232. Which integer matrices have integer inverses? ✓

Cramer's Rule

This is an application of the Cofactor Theorem to systems of linear equations. Consider a system

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\cdots \\ a_{n1}x_1 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

of n linear equations with n unknowns $(x_1, \ldots, x_n)$. It can be written in the matrix form

$$A\mathbf{x} = \mathbf{b},$$

where A is the $n \times n$-matrix of the coefficients a_{ij}, $\mathbf{b} = [b_1, \ldots, b_n]^t$ is the column of the right hand sides, and $\mathbf{x}$ is the column of unknowns. In the following Corollary, $\mathbf{a}_i$ denote columns of A.

`Corollary.` ***If*** $\det A$ ***is invertible then the system of linear equations*** $A\mathbf{x} = \mathbf{b}$ ***has a unique solution given by the***

formulas:

$$x_1 = \frac{\det[\mathbf{b}, \mathbf{a}_2, ..., \mathbf{a}_n]}{\det[\mathbf{a}_1, ..., \mathbf{a}_n]}, \ \ldots \ , x_n = \frac{\det[\mathbf{a}_1, ..., \mathbf{a}_{n-1}, \mathbf{b}]}{\det[\mathbf{a}_1, ..., \mathbf{a}_n]}.$$

Indeed, when $\det A \neq 0$, the matrix A is invertible. Multiplying the matrix equation $A\mathbf{x} = \mathbf{b}$ by A^{-1} on the left, we find: $\mathbf{x} = A^{-1}\mathbf{b}$. Thus the solution is unique, and $x_i = (\det A)^{-1}(A_{1i}b_1 + ... + A_{ni}b_n)$ according to the cofactor formula for the inverse matrix. But the sum $b_1A_{1i} + ... + b_nA_{ni}$ is the cofactor expansion for $\det[\mathbf{a}_1, ..., \mathbf{a}_{i-1}, \mathbf{b}, \mathbf{a}_{i+1}, ..., \mathbf{a}_n]$ with respect to the i-th column.

`Example.` Suppose that $a_{11}a_{22} \neq a_{12}a_{21}$. Then the system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_2 + a_{22}x_2 &= b_2 \end{aligned}$$

has a unique solution

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}, \quad x_2 = \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}.$$

EXERCISES

233. Solve systems of equations using Cramer's rule:

$$(a) \quad \begin{array}{rcl} 2x_1 - x_2 - x_3 &=& 4 \\ 3x_1 + 4x_2 - 2x_3 &=& 11 \\ 3x_1 - 2x_2 + 4x_3 &=& 11 \end{array}, \qquad (b) \quad \begin{array}{rcl} x_1 + 2x_2 + 4x_3 &=& 31 \\ 5x_1 + x_2 + 2x_3 &=& 29 \\ 3x_1 - x_2 + x_3 &=& 10 \end{array}. \qquad \checkmark$$

Three Cool Formulas

We collect here some useful generalizations of previous results.

A. We don't know of any reasonable generalization of determinants to the situation when matrix entries do *not* commute. However the following generalization of the formula $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$ is instrumental in some non-commutative applications.[10]

[10] Notably in the definition of *Berezinian* in super-mathematics [7].

In the block matrix $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$***, assume that*** D^{-1} ***exists. Then*** $\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A - BD^{-1}C) \det D$***.***

`Proof:` $\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} = \begin{bmatrix} A - BD^{-1}C & B \\ 0 & D \end{bmatrix}$.

B. Lagrange's formula[11] below generalizes cofactor expansions.

By a **multi-index** I of length $|I| = k$ we mean an increasing sequence $i_1 < \cdots < i_k$ of k indices from the set $\{1, \ldots, n\}$. Given and $n \times n$-matrix A and two multi-indices I, J of the same length k, we define the (IJ)**-minor** of A as the determinant of the $k \times k$-matrix formed by the entries $a_{\mathbf{i}_\alpha j_\beta}$ of A located at the intersections of the rows $i_1, \ldots, i_k$ with columns $j_1, \ldots, j_k$ (see Figure 30). Also, denote by $\bar{I}$ the multi-index **complementary** to I, i.e. formed by those $n - k$ indices from $\{1, \ldots, n\}$ which are *not* contained in I.

For each multi-index $I = (i_1, \ldots, i_k)$***, the following*** **cofactor expansion with respect to rows** $i_1, \ldots, i_k$ **holds true:**

$$\det A = \sum_{J : |J| = k} (-1)^{i_1 + \cdots + i_k + j_1 + \cdots + j_k} M_{IJ} M_{\bar{I}\bar{J}},$$

where the sum is taken over all multi-indices $J = (j_1, \ldots, j_k)$ ***of length*** k***.***

Similarly, one can similarly write Lagrange's cofactor expansion formula with respect to given k columns.

`Example.` Let $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$ and $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4$ be 8 vectors on the plane. Then $\begin{vmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 \\ \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{b}_4 \end{vmatrix} = |\mathbf{a}_1\ \mathbf{a}_2||\mathbf{b}_3\ \mathbf{b}_4| - |\mathbf{a}_1\ \mathbf{a}_3||\mathbf{b}_2\ \mathbf{b}_4|$ $+ |\mathbf{a}_1\ \mathbf{a}_4||\mathbf{b}_2\ \mathbf{b}_3| - |\mathbf{a}_2\ \mathbf{a}_3||\mathbf{b}_1\ \mathbf{b}_4| + |\mathbf{a}_2\ \mathbf{a}_4||\mathbf{b}_1\ \mathbf{b}_3| - |\mathbf{a}_3\ \mathbf{a}_4||\mathbf{b}_1\ \mathbf{b}_2|$.

[11] After Joseph-Louis **Lagrange** (1736–1813).

In the proof of Lagrange's formula, it suffices to assume that it is written with respect to the *first* k rows, i.e. that $I = (1, \dots, k)$. Indeed, interchanging them with the rows $i_1 < \cdots < i_k$ takes $(i_1 - 1) + (i_2 - 2) + \cdots + (i_k - k)$ transpositions, which is accounted for by the sign $(-1)^{i_1 + \cdots + i_k}$ in the formula.

Next, multiplying out $M_{IJ} M_{\bar{I}\bar{J}}$, we find $k!(n-k)!$ elementary products of the form:

$$\pm a_{1,j_{\alpha_1}} \cdots a_{k,j_{\alpha_k}} \; a_{k+1,\bar{j}_{\beta_1}} \cdots a_{n,\bar{j}_{\beta_{n-k}}},$$

where $\alpha = \begin{pmatrix} 1 & \dots & k \\ \alpha_1 & \dots & \alpha_k \end{pmatrix}$ and $\beta = \begin{pmatrix} 1 & \dots & n-k \\ \beta_1 & \dots & \beta_{n-k} \end{pmatrix}$ are permutations, and $j_{\alpha_\mu} \in J$, $\bar{j}_{\beta_\nu} \in \bar{J}$. It is clear that the total sum over multi-indices I contains each elementary product from $\det A$, and does it exactly once. Thus, to finish the proof, we need to compare the signs.

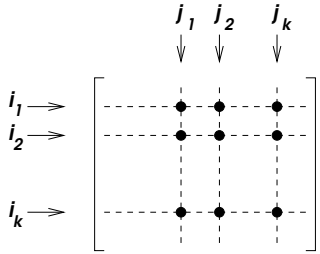

Figure 31

The sign $\pm$ in the above formula is equal to $\varepsilon(\alpha)\varepsilon(\beta)$, the product of the signs of the permutations α and β. The sign of this elementary product in the definition of $\det A$ is equal to the sign of the permutation $\begin{pmatrix} 1 & \dots & k & k+1 & \dots & n \\ j_{\alpha_1} & \dots & j_{\alpha_k} & \bar{j}_{\beta_1} & \dots & \bar{j}_{\beta_{n-k}} \end{pmatrix}$ on the set $J \cup \bar{J} = \{1, \dots, n\}$. Reordering separately the first k and last $n-k$ indices in the increasing order changes the sign of the permutation by $\varepsilon(\alpha)\varepsilon(\beta)$. Therefore the signs of all summands of $\det A$ which occur in $M_{IJ} M_{\bar{I}\bar{J}}$ are *coherent*. It remains to find the total sign with which $M_{IJ} M_{\bar{I}\bar{J}}$ occurs in $\det A$, by computing the sign of the permutation $\sigma := \begin{pmatrix} 1 & \dots & k & k+1 & \dots & n \\ j_1 & \dots & j_k & \bar{j}_1 & \dots & \bar{j}_{n-k} \end{pmatrix}$, where $j_1 < \cdots j_k$ and $\bar{j}_1 < \cdots < \bar{j}_{n-k}$.

Starting with the identity permutation $(1, 2 \dots, j_1, \dots, j_2, \dots, n)$, it takes $j_1 - 1$ transpositions of nearby indices to move j_1 to the 1st place. Then it takes $j_2 - 2$ such transpositions to move j_2 to the 2nd

place. Continuing this way, we find that

$$\varepsilon(\sigma) = (-1)^{(j_1-1)+\cdots+(j_k-k)} = (-1)^{1+\cdots+k+j_1+\cdots+\mathbf{j}_k}.$$

This agrees with Lagrange's formula, since $I = \{1, \dots, k\}$. $\square$.

C. Let A and B be $k \times n$ and $n \times k$ matrices (think of $k < n$). For each multi-index $I = (i_1, \dots, i_k)$, denote by A_I and B_I the $k \times k$-matrices formed by respectively: columns of A and rows of B with the indices $i_1, \dots, i_k$.

***The determinant of the $k \times k$-matrix AB is given by the following* Binet–Cauchy formula**:[12]

$$\det AB = \sum_I (\det A_I)(\det B_I).$$

Note that when $k = n$, this turns into the multiplicative property of determinants: $\det(AB) = (\det A)(\det B)$. Our second proof of it can be generalized to establish the formula of Binet–Cauchy. Namely, let $\mathbf{a}_1, \dots, \mathbf{a}_n$ denote columns of A. Then the jth column of $C = AB$ is the linear combination: $\mathbf{c}_j = \mathbf{a}_1 b_{1j} + \cdots + \mathbf{a}_n b_{nj}$. Using linearity in each $\mathbf{c}_j$, we find:

$$\det[\mathbf{c}_1, \dots, \mathbf{c}_k] = \sum_{1 \le i_1, \dots, i_k \le n} \det[\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_k}] b_{i_1 1} \cdots b_{i_k k}.$$

If any two of the indices i_α coincide, $\det[\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_k}] = 0$. Thus the sum is effectively taken over all *permutations* $\begin{pmatrix} 1 & \dots & k \\ i_1 & \dots & i_k \end{pmatrix}$ on the *set*[13] $\{i_1, \dots, i_k\}$. Reordering the columns $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_k}$ in the increasing order of the indices (and paying the "fees" ± 1 according to parities of permutations) we obtain the sum over all multi-indices of length k:

$$\sum_{i'_1 < \cdots < i'_k} \det[\mathbf{a}_{i'_1}, \dots, \mathbf{a}_{i'_k}] \sum_\sigma \varepsilon(\sigma) b_{i_1 1} \cdots b_{i_k k}.$$

The sum on the right is taken over permutations $\sigma = \begin{pmatrix} i'_1 & \dots & i'_k \\ i_1 & \dots & i_k \end{pmatrix}$. It is equal to $\det B_I$, where $I = (i'_1, \dots, i'_k)$. $\square$

`Corollary 1.` *If $k > n$, $\det AB = 0$.*

[12] After Jacques **Binet** (1786–1856) and Augustin Louis **Cauchy** (1789–1857).
[13] Remember that in a set, elements are unordered!

This is because no multi-indices of length $k > n$ can be formed from $\{1, \dots, n\}$. In the oppositely extreme case $k = 1$, Binet–Cauchy's formula turns into the expression $\mathbf{u}^t\mathbf{v} = \sum u_i v_i$ for the dot product of coordinate vectors. A "Pythagorean" interpretation of the following identity will come to light in the next chapter, in connection with volumes of parallelepipeds.

`Corollary 2.` $\det AA^t = \sum_I (\det A_I)^2$.

EXERCISES

234.* Compute determinants:

$$(a)\ \begin{vmatrix} 0 & x_1 & x_2 & \dots & x_n \\ x_1 & 1 & 0 & \dots & 0 \\ x_2 & 0 & 1 & \dots & 0 \\ . & . & . & . & . \\ x_n & 0 & \dots & 0 & 1 \end{vmatrix}, \quad (b)\ \begin{vmatrix} a & 0 & 0 & 0 & 0 & b \\ 0 & a & 0 & 0 & b & 0 \\ 0 & 0 & a & b & 0 & 0 \\ 0 & 0 & c & d & 0 & 0 \\ 0 & c & 0 & 0 & d & 0 \\ c & 0 & 0 & 0 & 0 & d \end{vmatrix} \qquad \text{ϟ ✓.}$$

235.* Let P_{ij}, $1 \le i < j \le 4$, denote the 2×2-minor of a 2×4-matrix formed by the columns i and j. Prove the following **Plücker identity**[14]

$$P_{12}P_{34} - P_{13}P_{24} + P_{14}P_{23} = 0. \qquad \checkmark$$

236. The **cross product** of two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ is defined by

$$\mathbf{x} \times \mathbf{y} := \left(\begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix}, \begin{vmatrix} x_3 & x_1 \\ y_3 & y_1 \end{vmatrix}, \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \right).$$

Prove that the length $|\mathbf{x} \times \mathbf{y}| = \sqrt{|\mathbf{x}|^2|\mathbf{y}|^2 - \langle \mathbf{x}, \mathbf{y} \rangle^2}$. ϟ

237.* Prove that $a_n + \cfrac{1}{a_{n-1} + \cfrac{1}{\dots + \cfrac{1}{a_1 + \cfrac{1}{a_0}}}} = \dfrac{\Delta_n}{\Delta_{n-1}}$,

where $\Delta_n = \begin{vmatrix} a_0 & 1 & 0 & \dots & 0 \\ -1 & a_1 & 1 & \dots & 0 \\ . & . & . & . & . \\ 0 & \dots & -1 & a_{n-1} & 1 \\ 0 & \dots & 0 & -1 & a_n \end{vmatrix}$. ϟ

238.* Compute: $\begin{vmatrix} \lambda & -1 & 0 & \dots & 0 \\ 0 & \lambda & -1 & \dots & 0 \\ . & . & . & . & . \\ 0 & \dots & 0 & \lambda & -1 \\ a_n & a_{n-1} & \dots & a_2 & \lambda + a_1 \end{vmatrix}$. ✓

[14] After Julius **Plücker** (1801–1868).

239.* Compute: $\begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \binom{2}{1} & \binom{3}{1} & \dots & \binom{n}{1} \\ 1 & \binom{3}{2} & \binom{4}{2} & \dots & \binom{n+1}{2} \\ . & . & . & . & . \\ 1 & \binom{n}{n-1} & \binom{n+1}{n-1} & \dots & \binom{2n-2}{n-1} \end{vmatrix}$. ↯ ✓

240.* Prove **Vandermonde's identity**[15]

$$\begin{vmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ . & . & . & . & . \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{vmatrix} = \prod_{1 \le i < j \le n} (x_j - x_i). \quad ↯$$

241.* Compute: $\begin{vmatrix} 1 & 2 & 3 & \dots & n \\ 1 & 2^3 & 3^3 & \dots & n^3 \\ . & . & . & . & . \\ 1 & 2^{2n-1} & 3^{2n-1} & \dots & n^{2n-1} \end{vmatrix}$. ↯ ✓

[15] After Alexandre-Theóphile **Vandermonde** (1735–1796).