Vectors and Matrices

2

2.1 Vectors

Ordered n-tuple of objects is called a vector

$$\mathbf{y} = (y_1, y_2, \ldots, y_n).$$

Throughout the text we confine ourselves to vectors the elements y_i of which are real numbers.

In contrast, a variable the value of which is a single number, not a vector, is called *scalar*.

Example 2.1. We can describe some economic unit **EU** by the vector

EU= (output, # of employees, capital stock, profit)

Given a vector $\mathbf{y} = (y_1, \ldots, y_n)$, elements y_i, $i = 1, \ldots, n$ are called *components* of the vector. We will usually denote vectors by bold letters.[1] The number n of components is called the *dimension* of the vector $\mathbf{y}$. The set of all n–dimensional vectors is denoted by $\mathbb{R}^n$ and called n-dimensional real space[2].

Two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are equal if $x_i = y_i$ for all $i = 1, 2, \ldots, n$.

Let $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ be two vectors. We compare these two vectors element by element and say that $\mathbf{x}$ is greater than $\mathbf{y}$ if for all i $x_i > y_i$, and denote this statement by $\mathbf{x} > \mathbf{y}$. Analogously, we can define $\mathbf{x} \geq \mathbf{y}$.

Note that, unlike in the case of real numbers, for vectors when $\mathbf{x} > \mathbf{y}$ does not hold, this does not imply $\mathbf{y} \geq \mathbf{x}$. Indeed, consider the vectors $\mathbf{x} = (1, 0)$ and $\mathbf{y} = (0, 1)$. It can be easily seen that neither $\mathbf{x} \geq \mathbf{y}$ nor $\mathbf{y} \geq \mathbf{x}$ is true.

[1]Some other notations for vectors are $\bar{y}$ and $\overrightarrow{y}$.

[2]The terms *arithmetic space, number space* and *coordinate space* are also used.

F. Aleskerov et al., *Linear Algebra for Economists*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-20570-5_2,

A vector $\mathbf{0} = (0, 0, \ldots, 0)$ (also denoted by $\bar{0}$) is called a *null vector*.[3]

A vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is called non-negative (which is denoted by $\mathbf{x} \geq \mathbf{0}$) if $x_i \geq 0$ for all i.

A vector $\mathbf{x}$ is called positive if $x_i > 0$ for all i. We denote this case by $\mathbf{x} > \mathbf{0}$.

2.1.1 Algebraic Properties of Vectors

One can define the following natural arithmetic operations with vectors.
Addition of two n-vectors

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$$

Subtraction of two n-vectors

$$\mathbf{x} - \mathbf{y} = (x_1 - y_1, x_2 - y_2, \ldots, x_n - y_n)$$

Multiplication of a vector by a real number λ

$$\lambda \mathbf{y} = (\lambda y_1, \lambda y_2, \ldots, \lambda y_n)$$

Example 2.2. Let $\mathbf{EU}_1 = (Y_1, L_1, K_1, P_1)$ be a vector representing an economic unit, say, a firm, see Example 2.1 (where, as usually, Y is its output, L is the number of employees, K is the capital stock, and P is the profit). Let us assume that it is merged with another firm represented by a vector $\mathbf{EU}_2 = (Y_2, L_2, K_2, P_2)$ (that is, we should consider two separate units as a single one). The resulting unit will be represented by a sum of two vectors

$$\mathbf{EU}_3 = (Y_1 + Y_2, L_1 + L_2, K_1 + K_2, P_1 + P_2) = \mathbf{EU}_1 + \mathbf{EU}_2.$$

In this situation, we have also $\mathbf{EU}_2 = \mathbf{EU}_3 - \mathbf{EU}_1$. Moreover, if the second firm is similar to the first one, we can assume that $\mathbf{EU}_1 = \mathbf{EU}_2$, hence the unit

$$\mathbf{EU}_3 = (2Y_1, 2L_1, 2K_1, 2P_1) = 2 \cdot \mathbf{EU}_1$$

gives also an example of the multiplication by a number 2.

This example, as well as other 'economic' examples in this book has an illustrative nature. Notice, however, that the profit of the merged firm might be higher or lower than the sum of two profits $P_1 + P_2$.

[3] The null vector is also called *zero vector*.

The following properties of the vector operations above follow from the definitions:

1a. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (commutativity).
1b. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (associativity).
1c. $\mathbf{x} + \mathbf{0} = \mathbf{x}$.
1d. $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.
2a. $1\mathbf{x} = \mathbf{x}$.
2b. $\lambda(\mu\mathbf{x}) = \lambda\mu(\mathbf{x})$.
3a. $(\lambda + \mu)\mathbf{x} = \lambda\mathbf{x} + \mu\mathbf{x}$.
3b. $\lambda(\mathbf{x} + \mathbf{y}) = \lambda\mathbf{x} + \lambda\mathbf{y}$.

Exercise 2.1. Try to prove these properties yourself.

2.1.2 Geometric Interpretation of Vectors and Operations on Them

Consider $\mathbb{R}^2$ plane. Vector $\mathbf{z} = (\alpha_1, \alpha_2)$ is represented by a directed line segment from the origin $(0, 0)$ to (α_1, α_2), see Fig. 2.1.

The sum of the two vectors $\mathbf{z}_1 = (\alpha_1, \beta_1)$ and $\mathbf{z}_2 = (\alpha_2, \beta_2)$ is obtained by adding up their coordinates, see Fig. 2.2.

In this figure, the sum $\mathbf{z}_1 + \mathbf{z}_2 = (\alpha_1 + \alpha_2, \beta_1 + \beta_2)$ is represented by a diagonal of a parallelogram sides of which being formed by the vectors $\mathbf{z}_1$ and $\mathbf{z}_2$.

Multiplication of a vector by a scalar has a contractionary (respectively, expansionary) effect if the scalar in absolute value is less (respectively, greater) than unity. The direction of the vector does not change if the scalar is positive, and it changes by 180 degrees if the scalar is negative. Figure 2.3 plots scalar multiplication for a vector $\mathbf{x}$, two scalars $\lambda_1 > 1$ and $-1 < \lambda_2 < 0$.

The difference of the two vectors $\mathbf{z}_2$ and $\mathbf{z}_1$ is shown on Fig. 2.4.

The projection of the vector $\mathbf{a}$ on $x-$axis is denoted by $pr_x\mathbf{a}$, and is shown in Fig. 2.5 below.

Let $\mathbf{z}_1, \ldots, \mathbf{z}_s$ be a set of vectors in $\mathbb{R}^n$. If there exist real numbers $\lambda_1, \ldots, \lambda_s$ not all being equal to 0 and

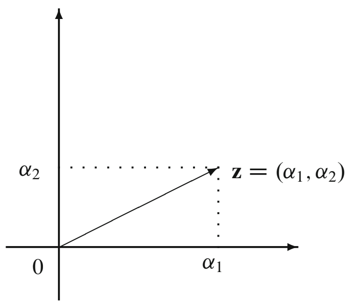


Fig. 2.1 A vector on the plane $\mathbb{R}^2$

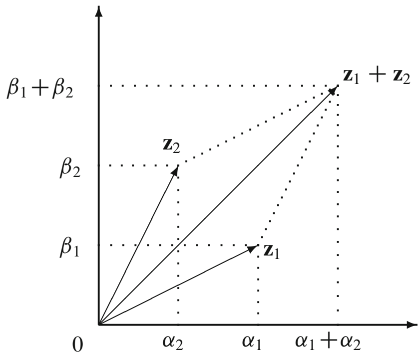


Fig. 2.2 The sum of two vectors

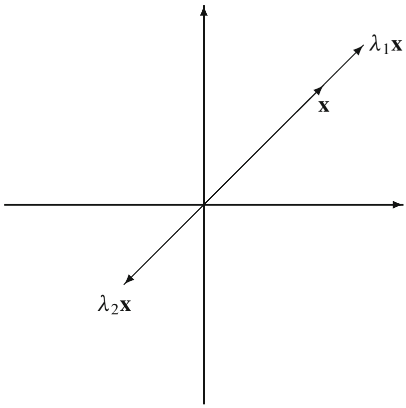


Fig. 2.3 The multiplication of a vector by a scalar

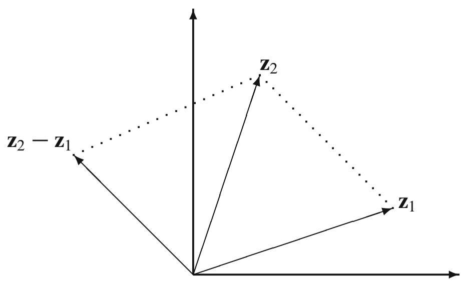


Fig. 2.4 The difference of vectors

$$\lambda_1\mathbf{z}_1+\lambda_2\mathbf{z}_2+\cdots+\lambda_s\mathbf{z}_s=\mathbf{0},$$

then these vectors are called *linearly dependent*.

Example 2.3. Three vectors $\mathbf{a}=(1,2,3)$, $\mathbf{b}=(4,5,6)$ and $\mathbf{c}=(7,8,9)$ are linearly dependent because

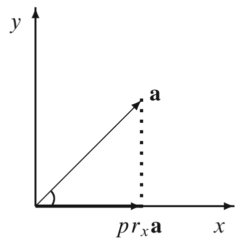


Fig. 2.5 The projection of a vector a on the x-axis

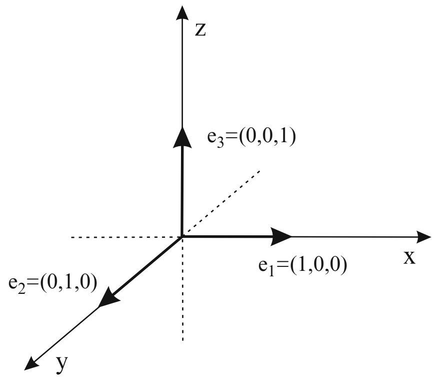


Fig. 2.6 Unit vectors in $\mathbb{R}^3$

$$1\mathbf{a} - 2\mathbf{b} + 1\mathbf{c} = \mathbf{0}.$$

The vectors $\mathbf{z}_1, \ldots, \mathbf{z}_s$ are called *linearly independent* if

$$\lambda_1 \mathbf{z}_1 + \cdots + \lambda_s \mathbf{z}_s = \mathbf{0}$$

holds only whenever $\lambda_1 = \lambda_2 = \cdots = \lambda_s = 0$.

Note that the n vectors $\mathbf{e}_1 = (1, 0, \ldots, 0)$, $\mathbf{e}_2 = (0, 1, \ldots, 0)$, ..., $\mathbf{e}_n = (0, 0, \ldots, 1)$ (see Fig. 2.6 for the case $n = 3$) are linearly independent in $\mathbb{R}^n$.

Assume that vectors $\mathbf{z}_1, \ldots, \mathbf{z}_s$ are linearly dependent, i.e., there exists at least one λ_i, where $1 \le i \le s$, such that $\lambda_i \ne 0$ and

$$\lambda_1 \mathbf{z}_1 + \lambda_2 \mathbf{z}_2 + \cdots + \lambda_i \mathbf{z}_i + \cdots + \lambda_s \mathbf{z}_s = \mathbf{0}.$$

Then

$$\lambda_i \mathbf{z}_i = -\lambda_1 \mathbf{z}_1 - \lambda_2 \mathbf{z}_2 - \cdots - \lambda_{i-1} \mathbf{z}_{i-1} - \lambda_{i+1} \mathbf{z}_{i+1} - \cdots - \lambda_s \mathbf{z}_s,$$

and

$$\mathbf{z}_i = \mu_1 \mathbf{z}_1 + \cdots + \mu_{i-1} \mathbf{z}_{i-1} + \mu_{i+1} \mathbf{z}_{i+1} + \cdots + \mu_s \mathbf{z}_s, \tag{2.1}$$

where $\mu_j = -\lambda_j / \lambda_i$, for all $j \ne i$ and $j \in \{1, \ldots, s\}$.

A vector **a** is called a *linear combination* of the vectors $\mathbf{b}_1, \dots, \mathbf{b}_n$ if it can be represented as

$$\mathbf{a} = \alpha_1 \mathbf{b}_1 + \cdots + \alpha_n \mathbf{b}_n,$$

where $\alpha_1, \dots, \alpha_n$ are real numbers. In particular, (2.1) shows that the vector $\mathbf{z}_i$ is a linear combination of the vectors $\mathbf{z}_1, \dots, \mathbf{z}_{i-1}, \mathbf{z}_{i+1}, \dots, \mathbf{z}_s$.

These results can be formulated as

Theorem 2.1. *If vectors $\mathbf{z}_1, \dots, \mathbf{z}_s$ are linearly dependent, then at least one of them is a linear combination of other vectors. Vectors one of which is a linear combination of others are linearly dependent.*

2.1.3 Geometric Interpretation in $\mathbb{R}^2$

Are the vectors $\mathbf{z}$ and $\lambda\mathbf{z}$ (see Fig. 2.7) linearly dependent?

Note from Fig. 2.8 that the vector $\lambda_1\mathbf{z}_1 + \lambda_2\mathbf{z}_2$ is a linear combination of the vectors $\mathbf{z}_1$ and $\mathbf{z}_2$. Any three vectors in $\mathbb{R}^2$ are linearly dependent!

Remark 2.1. Consider the following n vectors in $\mathbb{R}^n$.

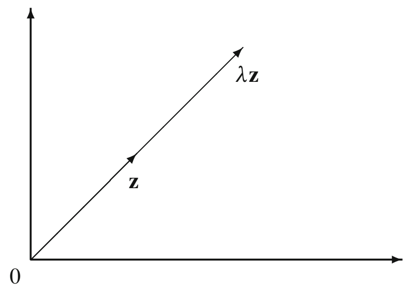


Fig. 2.7 Are these vectors linearly dependent?

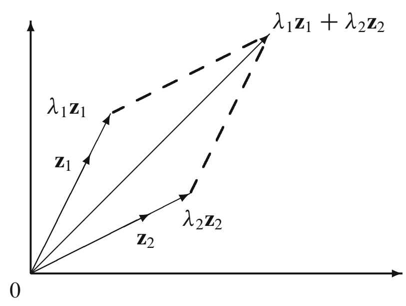


Fig. 2.8 A linear combination of two vectors

$$\begin{aligned}
\mathbf{a}_1 &= (1, -2, 0, 0, \dots, 0) \\
\mathbf{a}_2 &= (0, 1, -2, 0, \dots, 0) \\
&\vdots \\
\mathbf{a}_{n-1} &= (0, 0, \dots, 0, 1, -2) \\
\mathbf{a}_n &= (-2^{-(n-1)}, 0, \dots, 0, 0, 1)
\end{aligned}$$

These vectors are linearly dependent since

$$2^{-n}\mathbf{a}_1 + 2^{-(n-1)}\mathbf{a}_2 + \cdots + 2^{-1}\mathbf{a}_n = \mathbf{0}.$$

If $n > 40$ then $2^{-(n-1)} < 10^{-12}$, a very small number. Moreover, if $n > 64$, then $2^{-n} = 0$ for computers. So, for $n > 64$, we can assume that in our system $\mathbf{a}_n$ is given by $\mathbf{a}_n = (0, \dots, 0, 1)$. Thus, the system is written as

$$\begin{cases}
\mathbf{a}_1 = (1, -2, 0, 0, \dots, 0) \\
\mathbf{a}_2 = (0, 1, -2, 0, \dots, 0) \\
\vdots \\
\mathbf{a}_{n-1} = (0, 0, \dots, 0, 1, -2) \\
\mathbf{a}_n = (0, 0, \dots, 0, 0, \quad 1)
\end{cases}$$

But this system is linearly independent. (Check it!)

This example shows how sensitive might be linear dependency of vectors to rounding.

Exercise 2.2. Check if the following three vectors are linearly dependent:
(a) $\mathbf{a} = (1, 2, 1), \mathbf{b} = (-2, 3, -2), \mathbf{c} = (7, 4, 7)$;
(b) $\mathbf{a} = (1, 2, 3), \mathbf{b} = (0, -1, 3), \mathbf{c} = (2, -1, 2)$.

2.2 Dot Product of Two Vectors

Definition 2.1. For any two vectors $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$, the *dot product*[4] of $\mathbf{x}$ and $\mathbf{y}$ is denoted by $(\mathbf{x}, \mathbf{y})$, and is defined as

$$(\mathbf{x}, \mathbf{y}) = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = \sum_{i=1}^{n} x_i y_i \,. \tag{2.2}$$

[4] Other terms for dot product are *scalar product* and *inner product*.

Example 2.4. Let $\mathbf{a}_1 = (1, -2, 0, \dots, 0)$ and $\mathbf{a}_2 = (0, 1, -2, 0, \dots, 0)$. Then

$$(\mathbf{a}_1, \mathbf{a}_2) = 1 \cdot 0 + (-2) \cdot 1 + 0 \cdot (-2) + 0 \cdot 0 + \dots + 0 \cdot 0 = -2.$$

Example 2.5 (Household expenditures). Suppose the family consumes n goods. Let $\mathbf{p}$ be the vector of prices of these commodities (we assume competitive economy and take them as given), and $\mathbf{q}$ be the vector of the amounts of commodities consumed by this household. Then the total expenditure of the household can be obtained by dot product of these two vectors

$$\mathrm{E} = (\mathbf{p}, \mathbf{q}).$$

Dot product $(\mathbf{x}, \mathbf{y})$ of two vectors $\mathbf{x}$ and $\mathbf{y}$ is a real number and has the following properties, which can be checked directly:

1. $(\mathbf{x}, \mathbf{y}) = (\mathbf{y}, \mathbf{x})$ (symmetry or commutativity)
2. $(\lambda \mathbf{x}, \mathbf{y}) = \lambda(\mathbf{x}, \mathbf{y})$ for all $\lambda \in \mathbb{R}$ (associativity with respect to multiplication by a scalar)
3. $(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) = (\mathbf{x}_1, \mathbf{y}) + (\mathbf{x}_2, \mathbf{y})$ (distributivity)
4. $(\mathbf{x}, \mathbf{x}) \geq 0$ and $(\mathbf{x}, \mathbf{x}) = 0$ iff $\mathbf{x} = \mathbf{0}$ (non-negativity and non-degeneracy).

2.2.1 The Length of a Vector, and the Angle Between Two Vectors

Definition 2.2. The length of a vector $\mathbf{x}$ in $\mathbb{R}^n$ is defined as $\sqrt{(\mathbf{x}, \mathbf{x})}$ and denoted by $\mid \mathbf{x} \mid$. If $\mathbf{x} = (x_1, \dots, x_n)$ then $\mid \mathbf{x} \mid = \sqrt{x_1^2 + \dots + x_n^2}$. The angle φ between any two nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ is defined as

$$\cos \varphi = \frac{(\mathbf{x}, \mathbf{y})}{\mid \mathbf{x} \mid \mid \mathbf{y} \mid}, \quad 0 \leq \varphi \leq \pi. \tag{2.3}$$

We will see below that this definition of $\cos \varphi$ is correct, that is, the right hand side of the above formula belongs to the interval $[-1, 1]$.

Let us show first that the angle between two vectors $\mathbf{x}$ and $\mathbf{y}$ in the Cartesian plane is the geometric angle (Fig. 2.9).

Take any two vectors $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$ in $\mathbb{R}^2$. Then $\mathbf{y} - \mathbf{x} = (y_1 - x_1, y_2 - x_2)$. By the law of cosines we have

$$|\mathbf{y} - \mathbf{x}|^2 = |\mathbf{y}|^2 + |\mathbf{x}|^2 - 2\,|\mathbf{y}|\,|\mathbf{x}| \cos \varphi,$$

or

$$(y_1 - x_1)^2 + (y_2 - x_2)^2 = y_1^2 + x_1^2 + y_2^2 + x_2^2 - 2\sqrt{y_1^2 + y_2^2}\sqrt{x_1^2 + x_2^2} \cos \varphi.$$

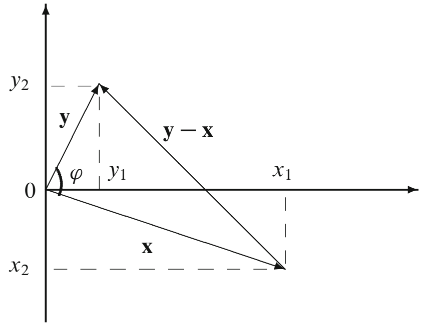


Fig. 2.9 The angle between two vectors

Then

$$\cos\varphi = \frac{y_1x_1 + y_2x_2}{\sqrt{y_1^2+y_2^2}\sqrt{x_1^2+x_2^2}} = \frac{(\mathbf{x},\mathbf{y})}{|\,\mathbf{x}\,||\,\mathbf{y}\,|}.$$

Definition 2.3. Two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ are called *orthogonal* (notation: $\mathbf{x}\perp\mathbf{y}$) if the angle between them is $\pi/2$, i.e. $(\mathbf{x},\mathbf{y})=0$.

Theorem 2.1 (Pythagoras). *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be two orthogonal vectors in* $\mathbb{R}^n$. *Then*

$$|\,\mathbf{x}+\mathbf{y}\,|^2 = |\,\mathbf{x}\,|^2 + |\,\mathbf{y}\,|^2\,. \tag{2.4}$$

Proof. $|\,\mathbf{x}+\mathbf{y}\,|^2 = (\mathbf{x}+\mathbf{y},\mathbf{x}+\mathbf{y}) = (\mathbf{x},\mathbf{x})+(\mathbf{x},\mathbf{y})+(\mathbf{y},\mathbf{x})+(\mathbf{y},\mathbf{y}) = |\,\mathbf{x}\,|^2+|\,\mathbf{y}\,|^2$ since $\mathbf{x}$ and $\mathbf{y}$ are orthogonal. □

The immediate generalization of the above theorem is the following one.

Theorem 2.2. Let $\mathbf{z}_1,\ldots,\mathbf{z}_s$ be a set of mutually orthogonal vectors in $\mathbb{R}^n$, i.e., for all i,j and $i\neq j$, $(\mathbf{z}_i,\mathbf{z}_j)=0$. Then

$$|\,\mathbf{z}_1+\mathbf{z}_2+\cdots+\mathbf{z}_s\,|^2 = |\,\mathbf{z}_1\,|^2+|\,\mathbf{z}_2\,|^2+\cdots+|\,\mathbf{z}_s\,|^2\,. \tag{2.5}$$

From the definition of the angle (2.3), it follows that

$$-1 \le \frac{(\mathbf{x},\mathbf{y})}{|\,\mathbf{x}\,||\,\mathbf{y}\,|} \le 1,$$

since $\varphi\in[0,\pi]$. The above inequalities can be rewritten as

$$\frac{(\mathbf{x},\mathbf{y})^2}{|\,\mathbf{x}\,|^2|\,\mathbf{y}\,|^2} \le 1,$$

or

$$(\mathbf{x}, \mathbf{y})^2 \leq (\mathbf{x}, \mathbf{x}) \cdot (\mathbf{y}, \mathbf{y}). \tag{2.6}$$

The inequality (2.6) is called *Cauchy*[5] *inequality.*

Let us prove it so that we can better understand why the angle φ between two vectors can take any value in the interval of $[0, \pi]$.

Proof. Given any two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, consider the vector $\mathbf{x} - \lambda\mathbf{y}$, where λ is a real number. By axiom 4 of dot product we must have

$$(\mathbf{x} - \lambda\mathbf{y}, \mathbf{x} - \lambda\mathbf{y}) \geq 0,$$

that is,

$$\lambda^2(\mathbf{y}, \mathbf{y}) - 2\lambda(\mathbf{x}, \mathbf{y}) + (\mathbf{x}, \mathbf{x}) \geq 0.$$

But then the discriminant of the quadratic equation

$$\lambda^2(\mathbf{y}, \mathbf{y}) - 2\lambda(\mathbf{x}, \mathbf{y}) + (\mathbf{x}, \mathbf{x}) = 0$$

can not be positive. Therefore, it must be true that

$$(\mathbf{x}, \mathbf{y})^2 - (\mathbf{x}, \mathbf{x}) \cdot (\mathbf{y}, \mathbf{y}) \leq 0.$$

□

Corollary 2.2. *For all* $\mathbf{x}$ *and* $\mathbf{y}$ *in* $\mathbb{R}^n$*,*

$$\mid \mathbf{x} + \mathbf{y} \mid \leq \mid \mathbf{x} \mid + \mid \mathbf{y} \mid . \tag{2.7}$$

Proof. Note that

$$\mid \mathbf{x} + \mathbf{y} \mid^2 = (\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) = (\mathbf{x}, \mathbf{x}) + 2(\mathbf{x}, \mathbf{y}) + (\mathbf{y}, \mathbf{y})$$

Now using $2(\mathbf{x}, \mathbf{y}) \leq 2 \mid (\mathbf{x}, \mathbf{y}) \mid \leq 2 \mid \mathbf{x} \mid\mid \mathbf{y} \mid$ by Cauchy inequality, we obtain

$$\begin{aligned} \mid \mathbf{x} + \mathbf{y} \mid^2 &\leq (\mathbf{x}, \mathbf{x}) + 2 \mid \mathbf{x} \mid \cdot \mid \mathbf{y} \mid + (\mathbf{y}, \mathbf{y}) \\ &= (\mid \mathbf{x} \mid + \mid \mathbf{y} \mid)^2 \end{aligned}$$

implying the desired result. □

[5] Augustin Louis Cauchy (1789–1857) was a great French mathematician. In addition to his works in algebra and determinants, he had created a modern approach to calculus, so-called epsilon–delta formalism.

Exercise 2.3. Plot the vectors $\mathbf{u} = (1, 2)$, $\mathbf{v} = (-3, 1)$ and their sum $\mathbf{w} = \mathbf{u} + \mathbf{v}$ and check visually the above inequality.

Exercise 2.4. Solve the system of equations

$$\begin{cases} (0, 0, 1, 1) \perp \mathbf{x}, \\ (1, 2, 0, -1) \perp \mathbf{x}, \\ \langle \mathbf{x}, \mathbf{a} \rangle = |\mathbf{a}| \cdot |\mathbf{x}|, \end{cases}$$

where $\mathbf{a} = (2, 1, 0, 0)$ and $\mathbf{x}$ is an unknown vector from $\mathbb{R}^4$.

Exercise 2.5. Two vectors $\mathbf{a}$ and $\mathbf{b}$ are called *parallel* if they are linearly dependent (notation: $\mathbf{a} \| \mathbf{b}$). Solve the system of equations

$$\begin{cases} (0, 0, -3, 4) \| \mathbf{x}, \\ |\mathbf{x}| = 15. \end{cases}$$

Exercise 2.6. Find the maximal angle of the triangle ABC, where $A = (0, 1, 2, 0)$, $B = (0, 1, 0, -1)$ and $C = (1, 0, 0, 1)$ are three points in $\mathbb{R}^4$.

Exercise 2.7. Given three points $\mathrm{A}(0, 1, 2, 3)$, $\mathrm{B}(1, -1, 1, -1)$ and $\mathrm{C}(1, 1, 0, 0)$ in $\mathbb{R}^4$, find the length of the median AM of the triangle ABC.

2.3 An Economic Example: Two Plants

Consider a firm operating two plants in two different locations. They both produce the same output (say, 10 units) using the same type of inputs. Although the amounts of inputs vary between the plants the output level is the same.

The firm management suspects that the production cost in Plant 2 is higher than in Plant 1. The following information was collected from the managers of these plants.

PLANT 1

Input	Price	Amount used
Input 1	3	9
Input 2	5	10
Input 3	7	8

PLANT 2

Input	Price	Amount used
Input 1	4	8
Input 2	7	12
Input 3	3	9

Question 1. Does this information confirm the suspicion of the firm management?

Answer. In order to answer this question one needs to calculate the cost function. Let w_{ij} denote the price of the ith input at the jth plant and x_{ij} denote the quantity of ith input used in production jth plant ($i = 1, 2, 3$ and $j = 1, 2$). Suppose both of these magnitudes are perfectly divisible, therefore can be represented by real numbers. The cost of production can be calculated by multiplying the amount of each input by its price and summing over all inputs.

This means price and quantity vectors (**p** and **q**) are defined on real space and inner product of these vectors are defined. In other words, both **p** and **q** are in the space $\mathbb{R}^3$. The cost function in this case can be written as an inner product of price and quantity vectors as

$$c = (\mathbf{w}, \mathbf{q}), \tag{2.8}$$

where c is the cost, a scalar. Using the data in the above tables cost of production can be calculated by using (2.8) as:

In Plant 1 the total cost is 133, which implies that unit cost is 13.3.

In Plant 2, on the other hand, cost of production is 143, which gives unit cost as 14.3 which is higher than the first plant.

That is, the suspicion is reasonable.

Question 2. The manager of the Plant 2 claims that the reason of the cost differences is the higher input prices in her region than in the other. Is the available information supports her claim?

Answer. Let the input price vectors for Plant 1 and 2 be denoted as $\mathbf{p}_1$ and $\mathbf{p}_2$. Suppose that the latter is a multiple λ of the former, i.e.,

$$\mathbf{p}_2 = \lambda \mathbf{p}_1.$$

Since both vectors are in the space $\mathbb{R}^3$, length is defined for both. From the definition of length one can obtain that

$$|\mathbf{p}_2| = \lambda |\mathbf{p}_1|.$$

In this case, however as can be seen from the tables this is not the case. Plant I enjoys lower prices for inputs 2 and 3, whereas Plant 2 enjoys lower price for input 3. For a rough guess, one can still compare the lengths of the input price vectors which are

$$|\mathbf{p}_1| = 9.11, |\mathbf{p}_2| = 8.60,$$

which indicates that price data does not support the claim of the manager of the Plant 2. When examined more closely, one can see that the Plant 2 uses the most expensive input (input 2) intensely. In contrast, Plant 2 managed to save from using the most expensive input (in this case input 3). Therefore, the manager needs to explain the reasons behind the choice mixture of inputs in her plant.

2.4 Another Economic Application: Index Numbers

One of the problems that applied economists deal with is how exactly the microeconomic information concerning many (in fact in millions) prices and quantities of goods can be aggregated into smaller number of price and quantity variables? Consider an economy which produces many different (in terms of quality, location and time) goods. This means there will thousands, if not millions, of prices to be considered.

Suppose, for example, one wants to estimate the rate of inflation for this economy. Inflation is the rate of change in the general price level, i.e., it has to be calculated by taking into account the changes in the prices of all goods. Assume that there are n different goods. Let p_i be the price and q_i is the quantity of the good i. Consider two points in time, 0 and t. Denote the aggregate value of all goods at time 0 and t, respectively, as

$$V^0 = \sum_i^n p_i^0 q_i^0 \tag{2.9}$$

and

$$V^t = \sum_i^n p_i^t q_i^t. \tag{2.10}$$

If $\mathbf{p}^0 = (p_1^0, \ldots, p_n^0)$ and $\mathbf{q}^0 = (q_1^0, \ldots, q_n^0)$ are the (row) vectors characterizing prices and quantities of goods, respectively, then $V^0 = (\mathbf{p}^0, \mathbf{q}^0)$ is just the dot product of vectors $\mathbf{p}^0$ and $\mathbf{q}^0$. Then V^t is the dot product of the vectors p^t and q^t, i.e. $V^t = (\mathbf{p}^t, \mathbf{q}^t)$.

Notice that, in general, between time 0 (initial period) and t (end period) both the prices and the quantities of goods vary. So simply dividing (2.10) by (2.9) will not give the rate of inflation. One needs to eliminate the effect of the change in the quantities. This is the index number problem which has a long history.[6]

In 1871, Laspeyres[7] proposed the following index number formula to deal with this problem

$$P_L = \frac{\sum_{i=1}^n p_i^t q_i^0}{\sum_{i=1}^n p_i^0 q_i^0} \tag{2.11}$$

Notice that in this formula prices are weighted by initial period quantity weights, in other words, Laspeyres assumed that price changes did not lead to a change in the composition of quantities.

---

[6]Charles de Ferrare Dutot is credited with the introduction of first price index in his book *Refléxions politiques sur les finances et le commerce* in 1738. He used the averages of prices, without weights.

[7]Ernst Louis Etienne Laspeyres (1834–1913) was a German economist and statistician, a representative of German historical school in economics.

In 1874, Paasche[8], suggested using end-period weights, instead of the initial period's

$$P_p = \frac{\sum_{i=1}^{n} p_i^t q_i^t}{\sum_{i=1}^{n} p_i^0 q_i^t}$$

Laspeyeres index underestimates, whereas Paasche index overestimates the actual inflation.

Exercise 2.8. Formulate Laspeyres and Paasche indices in term of price and quantity vectors.

Outline of the answer:

$$P_L = \frac{(\mathbf{p}^t, \mathbf{q}^0)}{(\mathbf{p}^0, \mathbf{q}^0)},$$

$$P_P = \frac{(\mathbf{p}^t, \mathbf{q}^t)}{(\mathbf{p}^0, \mathbf{q}^t)}.$$

Exercise 2.9. Consider a three good economy. The initial ($t = 0$) and end period's ($t = 1$) prices and quantities of goods are as given in the following table:

	Price ($t = 0$)	Quantity ($t = 0$)	Price ($t = 1$)	Quantity ($t = 1$)
Good 1	2	50	1,8	90
Good 2	1,5	90	2,2	70
Good 3	0,8	130	1	100

i. Estimate the inflation (i.e. percentage change in overall price level) for this economy by calculating Laspeyres index
ii. Repeat the same exercise by calculating Paasche index.

For further information on index numbers, we refer the reader to [9, 23].

2.5 Matrices

A matrix is a rectangular array of real numbers

$$\begin{bmatrix} a_{11} & a_{12} & . & . & . & a_{1n} \\ a_{21} & a_{22} & . & . & . & a_{2n} \\ \vdots & \vdots & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & & \vdots \\ \vdots & \vdots & & & \ddots & \vdots \\ a_{m1} & a_{m2} & . & . & . & a_{mn} \end{bmatrix}.$$

[8] Hermann Paasche (1851–1925), German economist and statistician, was a professor of political science at Aachen University.

We will denote matrices with capital letters A,B,... The generic element of a matrix A is denoted by a_{ij}, $i = 1, \dots, m$; $j = 1, \dots, n$, and the matrix itself is denoted briefly as $A = \|a_{ij}\|_{m \times n}$. Such a matrix with m rows and n columns is said to be of order $m \times n$ If the matrix is square (that is, $m = n$), it is simply said to be of order n.

We denote by $\mathbf{0}$ the *null* matrix which contains zeros only. The *identity* matrix is a matrix $I = I_n$ of size $n \times n$ whose elements are $i_{k,k} = 1$ and $i_{k,m} = 0$ for $k \neq m, k = 1, \dots, n$ and $m = 1, \dots, n$, that is, it has units on the diagonal and zeroes on the other places. The notion of the identity matrix will be discussed in Sect. 3.2.

Example 2.6. Object – property: Consider m economic units each of which is described by n indices. Units may be firms, and indices may involve the output, the number of employees, the capital stock, etc., of each firm.

Example 2.7. Consider an economy consisting of $m = n$ sectors, where for all $i, j \in \{1, 2, \dots, n\}$, a_{ij} denotes the share of the output produced in sector i and used by sector j, in the total output of sector i. (Note that in this case the row elements add up to one.)

Example 2.8. Consider $m = n$ cities. Here a_{ij} is the distance between city i and city j. Naturally, $a_{ii} = 0$, $a_{ij} > 0$, and $a_{ij} = a_{ji}$ for all $i \neq j$, and $i, j \in \{1, 2, \dots, n\}$.

We say that a matrix $A = \|a_{ij}\|_{m \times n}$ is non-negative if $a_{ij} \geq 0$ for all $i = 1, \dots, m$; $j = 1, \dots, n$. This case is simply denoted by $A \geq \mathbf{0}$.

Analogously is defined a positive matrix $A > \mathbf{0}$.

2.5.1 Operations on Matrices

Let $A = \|a_{ij}\|_{m \times n}$ and $B = \|b_{ij}\|_{m \times n}$ be two matrices. The sum of these matrices is defined as

$$A + B = \|a_{ij} + b_{ij}\|_{m \times n}.$$

Example 2.9.

$$\begin{bmatrix} 1 & 0 \\ 4 & 2 \\ 7 & 1 \end{bmatrix} + \begin{bmatrix} 3 & 2 \\ 7 & 3 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 11 & 5 \\ 11 & 2 \end{bmatrix}.$$

Let $A = \|a_{ij}\|_{m \times n}$ and $\lambda \in \mathbb{R}$. Then

$$\lambda A = \|\lambda a_{ij}\|_{m \times n}.$$

Example 2.10.

$$2\begin{bmatrix} 3 & 0 \\ 2 & 4 \\ 1 & 9 \end{bmatrix} = \begin{bmatrix} 6 & 0 \\ 4 & 8 \\ 2 & 18 \end{bmatrix}$$

Properties of Matrix Summation and Multiplication by a Scalar

(1-a) $A + B = B + A$.

(1-b) $A + (B + C) = (A + B) + C$.

(1-c) $A + (-A) = \mathbf{0}$, where $-A = (-1)A$.

(1-d) $A + \mathbf{0} = A$.

(2-a) $1A = A$.

(2-b) $\lambda(\mu A) = (\lambda\mu)A, \quad \lambda, \mu \in \mathbb{R}$.

(3-a) $0A = \mathbf{0}$.

(3-b) $(\lambda + \mu)A = \lambda A + \mu A, \quad \lambda, \mu \in \mathbb{R}$.

(3-c) $\lambda\,(A + B) = \lambda A + \lambda B, \quad \lambda, \mu \in \mathbb{R}$.

The properties of these two operations are the same as for vectors from $\mathbb{R}^n$. We will clarify this later in Chap. 6.

2.5.2 Matrix Multiplication

Let $A = \left\|a_{ij}\right\|_{m\times n}$ and $B = \left\|b_{jk}\right\|_{n\times p}$ be two matrices. Then the matrix AB of order $m \times p$ is defined as

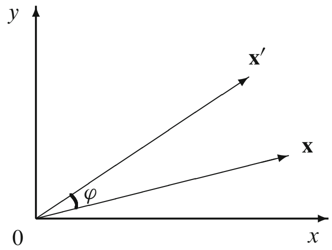
$$AB = \left[\sum_{j=1}^{n} a_{ij} b_{jk}\right]_{m\times p}$$

In other words, a product $C = AB$ of the above matrices A and B is a matrix $C = \left\|c_{ij}\right\|_{m\times p}$, where c_{ij} is equal to the dot product (A_i, B^j) of the i-th row A_i of the matrix A and the j-th column B^j of the matrix B considered as vectors from $\mathbb{R}^n$.

Consider 2×2 case. Given

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{ and } B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix},$$

Fig. 2.10 A rotation



we have

$$AB = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}.$$

Example 2.11.

$$\begin{bmatrix} 0 & 1 & 2 \\ 2 & 1 & 5 \end{bmatrix} \begin{bmatrix} 3 & 6 & 4 \\ 2 & 5 & 8 \\ 7 & 1 & 9 \end{bmatrix} = \begin{bmatrix} 16 & 7 & 26 \\ 43 & 22 & 61 \end{bmatrix}.$$

Example 2.12. Rotation of a vector $\mathbf{x} = (x, y)$ in $\mathbb{R}^2$ around the origin by a fixed angle φ (Fig. 2.10) can be expressed as a matrix multiplication. If $\mathbf{x}' = (x', y')$ is the rotated vector, then its coordinates can be expressed as

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = R_\alpha \begin{bmatrix} x \\ y \end{bmatrix}, \tag{2.12}$$

where

$$R_\alpha = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$$

is called a *rotation matrix*.

Note that if we consider the vectors $\mathbf{x}$ and $\mathbf{x}'$ as 1×2 matrices, then (2.12) may be briefly re-written as ${\mathbf{x}'}^T = R_\alpha \mathbf{x}^T$.

Exercise 2.10. Using elementary geometry and trigonometry, prove the equality (2.12).

Properties of Matrix Multiplication

(1-a) $\alpha(AB) = ((\alpha A)B) = A(\alpha B)$.
(1-b) $A(BC) = (AB)C$.
(1-c) $A\mathbf{0} = \mathbf{0}$.
(2-a) $A(B + C) = AB + AC$.
(2-b) $(A + B)C = AC + BC$.

Remark 2.2. Warning. $AB \neq BA$, in general.

Indeed, let A and B be matrices of order $m \times n$ and $n \times p$, respectively. To define the multiplication BA, we must have $p = m$. But matrices A and B may not commute even if both of them are square matrices of order $m \times m$. For example, consider

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \text{ and } B = \begin{bmatrix} -1 & 2 \\ 1 & 3 \end{bmatrix}.$$

We have

$$AB = \begin{bmatrix} 1 & 8 \\ 3 & 9 \end{bmatrix} \text{ while } BA = \begin{bmatrix} -1 & 4 \\ 1 & 11 \end{bmatrix}.$$

Exercise 2.11. Let A and B be square matrices such that $AB = BA$. Show that:
1. $(A + B)^2 = A^2 + 2AB + B^2$.
2. $A^2 - B^2 = (A - B)(A + B)$.

Exercise 2.12.* Prove the above properties of matrix multiplication.

Hint. To deduce the property 1-b), use the formula $\sum\limits_{i=1}^{n}\left(\sum\limits_{j=1}^{m} x_{ij}\right) = \sum\limits_{j=1}^{m}\left(\sum\limits_{i=1}^{n} x_{ij}\right)$.

Remark 2.3. The matrix multiplication defined above is one of the many concepts that are counted under the broader term "matrix product". It is certainly the most widely used one. However, there are two other matrix products that are of some interest to economists.

Kronecker Product of Matrices

Let $A = \|a_{ij}\|$ be an $m \times n$ matrix and $B = \|b_{ij}\|$ be a $p \times q$ matrix. Then the *Kronecker*[9] *product* of these two matrices is defined as

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

which is an $mp \times nq$ matrix. Kronecker product is also referred to as *direct product* or *tensor product* of matrices. For its use in econometrics, see [1, 8, 14].

[9] Leopold Kronecker (1823–1891) was a German mathematician who made a great contribution both to algebra and number theory. He was one of the founders of so-called constructive mathematics.

Hadamard Product of Matrices

The *Hadamard*[10] *product of matrices* (or *elementwise product*, or *Shur*[11] *product*) of two matrices $A = \|a_{ij}\|$ and $B = \|b_{ij}\|$ of the same dimensions $m \times n$ is a submatrix of the Kronecker product

$$A \circ B = \|a_{ij} b_{ij}\|_{m\times n}.$$

See [1, p. 340] and [24, Sect. 36] for the use of Hadamard product in matrix inequalities.

2.5.3 Trace of a Matrix

Given an $n \times n$ matrix $A = \|a_{ij}\|$, the sum of its diagonal elements $\mathrm{Tr}\, A = \sum_{i=1}^{n} a_{ii}$ is called the *trace* of the matrix A.

Example 2.13.

$$\mathrm{Tr} \begin{bmatrix} 1 & 2 & 3 \\ 10 & 20 & 30 \\ 100 & 200 & 300 \end{bmatrix} = 321$$

Exercise 2.13. Let A and B be two matrices of order n. Show that:

(a) $\mathrm{Tr}(A + B) = \mathrm{Tr}\, A + \mathrm{Tr}\, B$.
(b)* $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$.

2.6 Transpose of a Matrix

Let $A = \|a_{ij}\|_{m\times n}$. The matrix $B = \|b_{ij}\|_{n\times m}$ is called the *transpose* of A (and denoted by A^T) if $b_{ij} = a_{ji}$ for all $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$.

Example 2.14.

$$\begin{bmatrix} 3 & 0 \\ 2 & 4 \\ 1 & 9 \end{bmatrix}^T = \begin{bmatrix} 3 & 2 & 1 \\ 0 & 4 & 9 \end{bmatrix}$$

---

[10] Jacques Salomon Hadamard (1865–1963), a famous French mathematician who contributed in many branches of mathematics such as number theory, geometry, algebra, calculus and dynamical systems, as well as in optics, mechanics and geodesy. His most popular book *The psychology of invention in the mathematical field* (1945) gives a nice description of mathematical thinking.

[11] Issai Schur (1875–1941), an Israeli mathematician who was born in Belarus and died in Israel, made fundamental contributions to algebra, integral and algebraic equations, theory of matrices and number theory.

The transpose operator satisfies the following properties:

1. $(A^T)^T = A$.
2. $(A+B)^T = A^T + B^T$.
3. $(\alpha A)^T = \alpha A^T$.
4. $(AB)^T = B^T A^T$.

Proof. To prove the above properties, note that one can formally write

$$A^T = \left\|a_{ij}\right\|_{m\times n}^T = \left\|a_{ji}\right\|_{n\times m}.$$

Then $(A^T)^T = \left\|a_{ji}\right\|_{n\times m}^T = \left\|a_{ij}\right\|_{m\times n} = A$. This proves the property 1.

Now, $(A+B)^T = \left\|a_{ij}+b_{ij}\right\|_{m\times n}^T = \left\|a_{ji}+b_{ji}\right\|_{n\times m} = \left\|a_{ji}\right\|_{n\times m} + \left\|b_{ji}\right\|_{n\times m} = A^T + B^T$. This gives the second property.

To check the third one, we deduce that $(\alpha A)^T = \left\|\alpha a_{ij}\right\|_{m\times n}^T = \left\|\alpha a_{ji}\right\|_{n\times m}^T = \alpha\left\|a_{ji}\right\|_{n\times m}^T = \alpha A^T$.

Now, it remains to check the fourth property. Let $M = AB$ and $N = B^T A^T$, where the matrices A and B are of orders $m\times n$ and $n\times p$, respectively. Then $M = \left\|\alpha m_{ji}\right\|_{m\times p}$ with $m_{ij} = (A_i, B^j)$ and $N = \left\|\alpha n_{ij}\right\|_{p\times n}$ with $n_{ij} = ((B^T)_i, (A^T)^j)$. Since the transposition changes rows and columns, we have the equalities of vectors

$$(B^T)_i = B^i, (A^T)^j = A_j.$$

Hence, $m_{ij} = n_{ji}$ for all $i = 1,\ldots,m$ and $j = 1,\ldots,p$. Thus $M^T = N$, as desired. □

A matrix A is called *symmetric* if $A = A^T$. A simple example of symmetric matrices is the distance matrix $A = [a_{ij}]$, where a_{ij} is the distance between the cities i and j. Obviously, $a_{ij} = a_{ji}$ or $A = A^T$.

Theorem 2.3. *For each matrix A of order $n\times n$, the matrix AA^T is symmetric.*

Proof. Consider $(AA^T)^T$. By the properties 3) and 4), we have $(AA^T)^T = (A^T)^T A^T = AA^T$. □

Exercise 2.14. Let A and B be two matrices of order n. Show that $\mathrm{Tr}\, A^T = \mathrm{Tr}\, A$.

2.7 Rank of a Matrix

Are the vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ in the Fig. 2.11 linearly dependent?
It is obvious that there exists γ such that

$$\mathbf{u} + \gamma\mathbf{z} = \mathbf{0},$$

or

$$\alpha\mathbf{x} + \beta\mathbf{y} + \gamma\mathbf{z} = \mathbf{0},$$

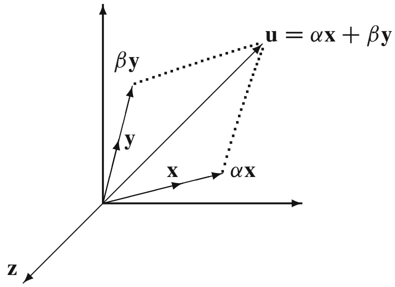


Fig. 2.11 Are the vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ linearly dependent?

i.e., these three vectors are linearly dependent.

Let us recall the notion of linear dependence of vectors. Consider vectors

$$\begin{aligned}\alpha &= (2, -5, 1, -1)\\ \beta &= (1, 3, 6, 5)\\ \gamma &= (-1, 4, 1, 2)\,.\end{aligned}$$

Are they linearly dependent? To answer, we construct a system of linear equations as follows: suppose the above vectors are linearly dependent. Then

$$a\alpha + b\beta + c\gamma = \mathbf{0}$$

for some parameters a, b, c, which are not all zero. In component-wise form, we obtain a homogeneous system of linear equation:

$$\begin{cases} 2a + 1b - c = 0\\ -5a + 3b + 4c = 0\\ a + 6b + c = 0\\ -a + 5b + 2c = 0\end{cases}$$

Here the system of linear equations is called *homogeneous* if every equation has the form “a linear combination of variables is equal to zero”.

One can check directly that a solution of the above system is given by $a = 7, b = -3$ and $c = 11$. Hence

$$7\alpha - 3\beta + 11\gamma = \mathbf{0}.$$

Consider now a matrix A

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n}\\ a_{21} & a_{22} & \dots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{s1} & a_{s2} & \dots & a_{sn}\end{bmatrix}.$$

Columns of this matrix can be considered as s-dimensional vectors, and maximal number of linearly independent columns is called the *rank* of A.

Example 2.15. Consider the matrix A with columns being the above vectors α, β and γ

$$A = \begin{bmatrix} 2 & 1 & -1 \\ -5 & 3 & 4 \\ 1 & 6 & 1 \\ -1 & 5 & 2 \end{bmatrix}.$$

Since A has 3 columns and the columns are linearly dependent, we have rank $A \leq 2$. On the other hand, it is easy to see that the first two columns of A are linearly independent, hence rank $A \geq 2$. Thus we conclude that rank $A = 2$.

Example 2.16. For the null matrix $\mathbf{0}$, we have the rank $A = 0$. On the other hand, the unit matrix I of the order $n \times n$ has the rank n.

Theorem 2.4. *The maximal number of linearly independent rows of a matrix equals to the maximal number of its linearly independent columns. Recalling the notion of the transpose, we have*

$$\text{rank}\, A = \text{rank}\, A^T$$

for every matrix A.

The proof of this theorem is given in Corollary 4.6.

Exercise 2.15. Check this statement for the above matrix A.

2.8 Elementary Operations and Elementary Matrices

In this section, we give a method to find linear dependence of columns of a matrix, and hence, to calculate its rank.

Let A be a matrix of order $m \times n$. Recall that its rows are n–vectors denoted by $A_1, A_2, \ldots, A_m$. The following simple transformations of A are called *elementary (row) operations*. All of them transform A to another matrix A' of the same order one or two rows (say, i-th and j-th) of which slightly differs from those of A:

1. Row switching: $A'_i = A_j$, $A'_j = A_i$.
2. Row multiplication: $A'_i = \lambda A_i$, where $\lambda \neq 0$ is a number.
3. Row replacement: $A'_i = A_i + \lambda A_j$, where $\lambda \neq 0$ is a number.

Example 2.17. Let us apply these operations to the unit matrix

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

The resulting matrices are called *elementary transformation matrices*; they are:

1. $$T_{i,j} = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \dots & 1 & & \\ & & \vdots & \ddots & \vdots & & \\ & & 1 & \dots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix};$$

2. $$T_i(\lambda) = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & \lambda & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix};$$

3. $$T_{i,j}(\lambda) = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \dots & \lambda & & \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}.$$

Exercise 2.16.* Show that any elementary operation of the second type is a composition of several operations of the first and the third type.

Theorem 2.5. *If A' is a result of an elementary operation of a matrix A, then*

$$A' = TA,$$

where T is a matrix of elementary transformation corresponding to the operation.

Exercise 2.17. Prove this Theorem 2.5.

(*Hint:* Use the definition of product of two matrices as a matrix entries of which are the dot products of rows and columns of the multipliers.)

Let t_1 and t_2 be elementary operations with corresponding matrices T_1 and T_2. The composition $t = t_1 t_2$ of these two operation is another (non-elementary) operation. It follows from Theorem 2.5 that t transforms any matrix A to a matrix

$$A' = t(A) = t_1(t_2(A)) = TA, \text{ where } T = T_1 T_2.$$

So, a matrix T corresponding to a composition of elementary operations is a product of matrices corresponding to the composers.

Another property of elementary operations is that all of them are invertible. This means that one can define the inverse operations for elementary operations of all three kinds listed above. For an operation of the first kind (switching), this inverse operation is the same switching; for the row multiplication by λ, it is a multiplication of the same row by $1/\lambda$; finally, for the replacement of A_i by $A'_i = A_i + \lambda A_j$ the inverse is a replacement of A'_i by $A'_i - \lambda A'_j = A_i$. Obviously, all these inverses are again elementary operations.

We obtain the following

Lemma 2.6. *Suppose that some elementary operation transforms a matrix A to A'. Then there is another elementary operation, which transforms the matrix A' to A.*

Another property of elementary operations is given in the following theorem.

Theorem 2.7. *Suppose that some columns $A^{i_1}, \dots, A^{i_k}$ of a matrix A are linearly dependent, that is, their linear combination is equal to zero*

$$\alpha_1 A^{i_1} + \dots + \alpha_k A^{i_k} = \mathbf{0}.$$

Let B be a matrix obtained from A by a sequence of several elementary operations. Then the corresponding linear combination of columns of B is also equal to zero

$$\alpha_1 B^{i_1} + \dots + \alpha_k B^{i_k} = \mathbf{0}.$$

Proof. Let $T_1, \dots, T_q$ be the matrices of elementary operations whose compositions transforms A to B. Then $B = TA$, where T is a matrix product $T = T_q \dots T_2 T_1$. This means that every column B^j of the matrix B is equal to TA^j. Thus,

$$\alpha_1 B^{i_1} + \dots + \alpha_k B^{i_k} = \alpha_1 TA^{i_1} + \dots + \alpha_k TA^{i_k} = T(\alpha_1 A^{i_1} + \dots + \alpha_k A^{i_k}) = T\mathbf{0} = \mathbf{0}.$$

$\square$

Corollary 2.8. *Let a matrix B be obtained from a matrix A by a sequence of several elementary operations. Then a collection $A^{i_1}, \dots, A^{i_k}$ of columns of the matrix A is linearly dependent if and only if corresponding collection $B^{i_1}, \dots, B^{i_k}$ is linearly dependent.*

In particular, this means that $\operatorname{rank} A = \operatorname{rank} B$.

Proof. The 'only if' statement immediately follows from Theorem 2.7.

According to Lemma 2.6, the matrix A as well may be obtained from B via a sequence of elementary operations (inverses of the given ones). Thus, we can apply the 'only if' part to the collection $B^{i_1}, \dots, B^{i_k}$ of columns of the matrix B. This imply the 'if' part.

By the definition of rank, the equality $\operatorname{rank} A = \operatorname{rank} B$ follows. □

Example 2.18. Let us find the rank of the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

Before calculations, we apply some elementary operations. First, let us substitute the third row: A_3 with $A_3 - 2A_2$. We get the matrix

$$A' = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ -1 & -2 & -3 \end{bmatrix}.$$

Now, substitute again: $A'_3 \mapsto A'_3 + A'_1$ and then $A'_2 \mapsto A'_2 - 4A'_1$. We obtain the matrix

$$A'' = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{bmatrix}.$$

Finally, let us substitute $A''_1 \to A''_1 + (2/3)A''_2$ and multiply $A''_2 \to (-1/3)A''_2$. We obtain the matrix

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

It is obvious that the first two columns of this matrix B are linearly independent while $B^3 = -B^1 + 2B^2$. Hence $\operatorname{rank} A = \operatorname{rank} B = 2$.

Definition 2.4. A matrix A is said to have a (row) *canonical form* (see Fig. 2.12), if the following four conditions are satisfied:

1. All nonzero rows are above any rows of all zeroes.
2. The first nonzero coefficient of any row (called also *leading coefficient*) is always placed to the right of the leading coefficient of the row above it.

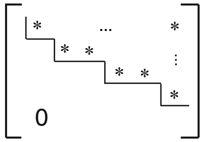


Fig. 2.12 Row echelon form of a matrix

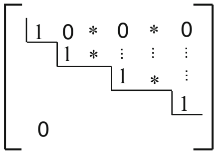


Fig. 2.13 Canonical form of a matrix

3. All leading coefficients are equal to 1.
4. All entries above a leading coefficient in the same column are equal to 0.

If only first two of the above conditions are satisfied, then the matrix is said to have a *row echelon form* (see Fig. 2.13).

Example 2.19. In Example 2.18 above, the marix A'' has a row echelon form while the matrix B has even a canonical form.

It is easy to calculate the rank of a matrix in an echelon form: it is simply equal to the number of nonzero rows in it.

Theorem 2.9. *Every matrix A can be transformed via a number of elementary operations to another matrix B in a row echelon form (and even in a canonical form). Then the rank of the matrix A is equal to the number of nonzero rows of the matrix B.*

Let us give an algorithm to construct an echelon form of the matrix. This algorithm is called the *Gaussian*[12] *elimination procedure*. It reduces all columns of the matrix one-by-one to the columns of some matrix in a row echelon form. In a recent step, we assume that a submatrix consisting of the first $(j-1)$ columns has an echelon form. Suppose that this submatrix has $(i-1)$ nonzero rows.

In the j-th step, we provide the following:

1. If all elements of the j-th column beginning with a_{ij} and below are equal to zero, the procedure is terminated. Then we go to the $(j+1)$-th step of the algorithm.
2. Otherwise, find the first nonzero element (say, a_{ij}) in the j-th column in the i-th row and below. If it is not a_{ij}, switch two rows A_i and A_j of the matrix. (see Fig. 2.14).

   Now, we obtain a matrix A such that $a_{pk} \neq 0$ (Fig. 2.15).

[12] Carl Friedrich Gauss (1777–1855) was a great German mathematician and physicist. He made fundamental contribution to a lot of branches of pure and applied mathematics including geodesy, statistics, and astronomy.

Fig. 2.14 The Gaussian elimination, row switching

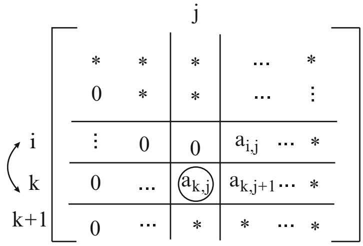


Fig. 2.15 The Gaussian elimination, row subtraction

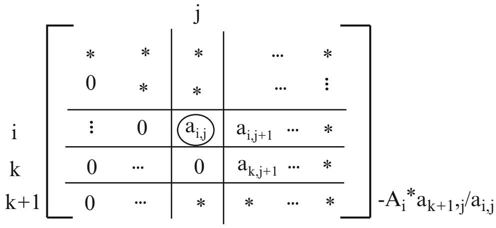
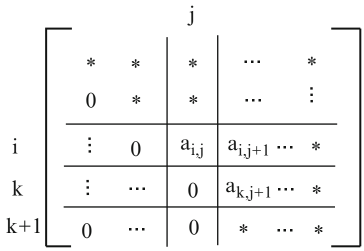


Fig. 2.16 The Gaussian elimination, the result of row subtractions



3. For every $p > i$, provide the following row replacement: the row $A_p \to A_p - (a_{pj}/a_{ij})A_j$ (Fig. 2.16).

These three types of operations are sufficient to construct a row echelon form. In the next step of the algorithm, we take $p + 1$ in place of p and $j + 1$ in place of j.

Note that in the above Example 2.18 we used this algorithm to transform A' to A''. Another example is given below.

Example 2.20. Using Gauss algorithm, let us construct a row echelon form of the matrix

$$A = \begin{bmatrix} 0 & 0 & 3 \\ 2 & 6 & -2 \\ 4 & 12 & -1 \end{bmatrix}.$$

In the beginning, $i = j = 1$, that is, we begin with the first column. In operation 1, we find the first nonzero element of this column, that is, a_{21}. In operation 2, we switch the first and the second rows and get the matrix

$$\begin{bmatrix} 2 & 6 & -2 \\ 0 & 0 & 3 \\ 4 & 12 & -1 \end{bmatrix}.$$

In operation 3, we subtract the doubled first row from the third one and get the matrix

$$\begin{bmatrix} 2 & 6 & -2 \\ 0 & 0 & 3 \\ 0 & 0 & 3 \end{bmatrix}.$$

Now, we provide the same three steps for the submatrix formed by the last two rows. The first nonzero column of the submatrix is the third one, so, in operation 1 we put $p = 3$. In operation 2, we find the first nonzero element of the column of the submatrix ($a_{23} = 3$). In operation 3, we replace the first row A_1 by $A_1 + (1/3)A_2$ and the third row A_3 by $A_3 - A_2$. We obtain a matrix in a row echelon form

$$\begin{bmatrix} 2 & 6 & -2 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix}.$$

The next theorem gives a stronger version of Gaussian elimination.

Theorem 2.10. *1. Every matrix A can also be transformed via elementary operations to a matrix C in a canonical form.*

2. The above canonical form C is unique for every matrix A, that is, it does not depend on the sequence of elementary operations which leads to this from.

Exercise 2.18. Prove the above Theorem 2.10.

Hint. To prove the first part, extend the above algorithm in the following way. To construct a canonical form, we need the same operations 1 and 2 as in the Gauss algorithm, a modified version of the above operation 3 and an additional operation 4.

(3') For every $p \neq i$, provide the following row replacement: the row $A_p \to A_p - (a_{pj}/a_{ij})A_j$.

(4) Replace the i-th row A_i by $(1/a_{ij})A_i$, that is, divide the i-th row by its first nonzero coefficient a_{ij}.

For the second part of the theorem, use Corollary 2.8.

Exercise 2.19. Find the canonical form of the matrix from Example 2.20.

2.9 Problems

1. Find a vector $\mathbf{x}$ such that:
   (a) $\mathbf{x} + \mathbf{y} = \mathbf{z}$, where $\mathbf{y} = (0, 3, 4, -2)$, and $\mathbf{z} = (3, 2, 1, -5)$.
   (b) $5\mathbf{x} = \mathbf{y} - \mathbf{z}$, where $\mathbf{y} = (-1, -1, 2)$ and $\mathbf{z} = (0, 1, 7)$.

2. Let $\mathbf{x}$ and $\mathbf{y}$ be two vectors in $\mathbb{R}^n$. Prove that:
   (a) $\mathbf{x} + \mathbf{y} = \mathbf{x}$ if and only if $\mathbf{y} = \mathbf{0}$.
   (b) $\lambda \mathbf{x} = \mathbf{0}$ and $\lambda \neq 0$ if and only if $\mathbf{x} = \mathbf{0}$.
3. Prove that vectors $\mathbf{z}_1, \ldots, \mathbf{z}_s$ in $\mathbb{R}^n$ are linearly dependent if one of them is the null vector.
4. Are the vectors below linearly dependent?

$$\begin{aligned}\mathbf{a}_1 &= (1,\ \ 0,\ 0,\ 2,\ \ 5)\\ \mathbf{a}_2 &= (0,\ \ 1,\ 0,\ 3,\ \ 4)\\ \mathbf{a}_3 &= (0,\ \ 0,\ 1,\ 4,\ \ 7)\\ \mathbf{a}_4 &= (2, -3, 4, 11, 12)\end{aligned}$$

5. Let $\mathbf{z}_1, \ldots, \mathbf{z}_s$ be linearly independent vectors and $\mathbf{x}$ be a vector such that

$$\mathbf{x} = \lambda_1 \mathbf{z}_1 + \cdots + \lambda_s \mathbf{z}_s,$$

   where $\lambda_i \in \mathbb{R}$ for all i. Show that this representation is unique.
6. Show that n vectors given by

$$\begin{aligned}\mathbf{x}_1 &= (1, 0, 0, \ldots 0, 0)\\ \mathbf{x}_2 &= (0, 1, 0, \ldots 0, 0)\\ &\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\\ \mathbf{x}_n &= (0, 0, 0, \ldots 0, 1)\end{aligned}$$

   are linearly independent in $\mathbb{R}^n$.
7. Find the rank of the following matrices:

$$\text{(a)} \begin{bmatrix} 2 & -1 & 3 & -2 & 4 \\ 4 & -2 & 5 & 1 & 7 \\ 2 & -1 & 1 & 8 & 2 \end{bmatrix}; \text{(b)} \begin{bmatrix} 3 & -1 & 3 & 2 & 5 \\ 5 & -3 & 2 & 3 & 4 \\ 1 & -3 & -5 & 0 & -7 \\ 7 & -5 & 1 & 4 & 1 \end{bmatrix}.$$

8. Show that n vectors given by

$$\begin{aligned}\mathbf{x}_1 &= (\eta_{11}, \eta_{12}, \ldots \eta_{1,n-1}, \eta_{1n})\\ \mathbf{x}_2 &= (0, \eta_{22}, \ldots \eta_{2,n-1}, \eta_{2n})\\ &\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\\ \mathbf{x}_n &= (0, 0, \ldots 0, \eta_{nn})\end{aligned}$$

   are linearly independent in $\mathbb{R}^n$ if $\eta_{ii} \neq 0$ for all i.
9. Check that in Definition 2.1 all axioms $1 - 4$ are satisfied.
10. Show that the Cauchy inequality (2.6) holds with the equality sign if $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent.
11. How many boolean (with components equal to 0 or 1) vectors exist in $\mathbb{R}^n$?
12. Find an example of matrices A, B and C such that $AB = AC$, $A \neq \mathbf{0}$, and $B \neq C$.
13. Find an example of matrices A and B such that $A \neq \mathbf{0}$, $B \neq \mathbf{0}$, but $AB = \mathbf{0}$.
14. Show that $A\mathbf{0} = \mathbf{0}A = \mathbf{0}$.

15. Prove that $(\alpha A)(\beta B) = (\alpha\beta)\,AB$ for all real numbers α and β, and for all matrices A and B such that the matrix products exist.
16. Prove that $(\alpha A)\,B = \alpha\,(AB) = A\,(\alpha B)$ for each real number α and for all matrices A and B such that the matrix products exist.
17. Let A, B and C be $n \times n$ matrices. Show that $ABC = CAB$ if $AC = CA$ and $BC = CB$.
18. Find a 2×3 matrix A and a 3×2 matrix B such that

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

19. Let

$$A = \begin{bmatrix} 1 & -1 \\ -3 & 3 \end{bmatrix}.$$

(a) Find $\mathbf{x} \neq \mathbf{0}$ such that $A\mathbf{x} = \mathbf{0}$.
(b) Find $\mathbf{y} \neq \mathbf{0}$ such that $\mathbf{y}A = \mathbf{0}$.

20. Let α and β be two angles. Prove the following property of rotation matrices:

$$R_{\alpha+\beta} = R_\alpha R_\beta.$$

21. Prove the properties of matrix summation.
22. Calculate

$$\begin{bmatrix} 0 & 2 & -1 \\ -2 & -1 & 2 \\ 3 & -2 & -1 \end{bmatrix} \begin{bmatrix} 70 & 34 & -107 \\ 52 & 26 & -68 \\ 101 & 50 & -140 \end{bmatrix} \begin{bmatrix} 27 & -18 & 10 \\ -46 & 31 & -17 \\ 3 & 2 & 1 \end{bmatrix}.$$

23. How $A \cdot B$ will change if:
(a) ith and jth rows of A are interchanged?
(b) a constant c times jth row of A is added to its ith row?
(c) ith and jth columns of B are interchanged?
(d) a constant c times jth column of B is added to its ith column?

24.* Show that $\operatorname{rank}(AB) \leq \operatorname{rank} A$ and $\operatorname{rank}(AB) \leq \operatorname{rank} B$.

25.* Show that the sum of the entries of the Hadamard product $A \circ B$ of two matrices A and B of order n (so-called a *Frobenius*[13] *product*) $(A, B)_F$ is equal to $\operatorname{Tr} AB^T$.

26.* Prove that any matrix A can be represented as $A = B + C$, where B is symmetric matrix and C is an anti-symmetric matrix (i.e., $C^T = -C$).

27. Find all 2×2 matrices A satisfying $A^2 = \mathbf{0}$.
28. Find all 2×2 matrices A satisfying $A^2 = I_2$.

[13] Ferdinand Georg Frobenius (1849–1917) was a famous German algebraist. He made a great contribution to group theory and also proved a number of significant theorems in algebraic equations, geometry, number theory, and theory of matrices.

29. Find a row echelon form and the rank of the matrix

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

30. Find the canonical form of the matrix

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 5 & 15 & 2 & 1 \\ -2 & -6 & 1 & 3 \end{bmatrix}.$$

http://www.springer.com/978-3-642-20569-9

Linear Algebra for Economists
Aleskerov, F.; Ersel, H.; Piontkovski, D.
2011, XII, 280 p., Hardcover
ISBN: 978-3-642-20569-9