Matrix

Inner product and geometry

Gyula Magyarkuti

October 6, 2024

Department of Mathematics
Corvinus University of Budapest
S208/B

email: magyarkuti@uni-corvinus.hu
web: https://magyarkuti.github.io/linear-algebra

Outline

Inner product

Concept of Matrices

Matrix operations

Multiplication of matrices

Powers of matrix

The rank

Inner product

INNER PRODUCT IN $\mathbb{R}^n$

To generalize the geometric concepts known in the plane or 3-dimensional space some new notion must be introduced.

Definition

The inner product of two vectors $\mathbf{a} = (a_1, ..., a_n)$ and $\mathbf{b} = (b_1, ..., b_n) \in \mathbb{R}^n$ is the real number

$$\langle \mathbf{a}, \mathbf{b} \rangle = a_1 b_1 + ... + a_n b_n.$$

Theorem (Properties of inner product)

1. $\forall \mathbf{a}, \mathbf{b} \in \mathbb{R}^n : \langle \mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{b}, \mathbf{a} \rangle$,
2. $\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n : \langle \mathbf{a} + \mathbf{b}, \mathbf{c} \rangle = \langle \mathbf{a}, \mathbf{c} \rangle + \langle \mathbf{b}, \mathbf{c} \rangle$,
3. $\forall \mathbf{a}, \mathbf{b} \in \mathbb{R}^n, \forall \lambda \in \mathbb{R} : \langle \lambda \mathbf{a}, \mathbf{b} \rangle = \lambda \langle \mathbf{a}, \mathbf{b} \rangle$,
4. $\langle \mathbf{a}, \mathbf{a} \rangle \geq 0$ *and* $\langle \mathbf{a}, \mathbf{a} \rangle = 0$, *iff* $\mathbf{a} = \mathbf{0}$.

Example

Let $\mathbf{a} = (3, -6, 1)$ and $\mathbf{b} = (1, 0, 4) \in \mathbb{R}^3$. Find the inner product $\langle 2\mathbf{a} - 3\mathbf{b}, \mathbf{a} + 2\mathbf{b} \rangle$.

$2\mathbf{a} - 3\mathbf{b} =(3, -12, -10)$ and $\mathbf{a} + 2\mathbf{b} =(5, -6, 9)$ thus their inner product

$$\langle 2\mathbf{a} - 3\mathbf{b}, \mathbf{a} + 2\mathbf{b} \rangle = 3 \cdot 5 + (-12) \cdot (-6) + (-10) \cdot 9 = -3.$$

another solution can be obtained if we use the properties of the inner product:

$$\begin{aligned} \langle 2\mathbf{a} - 3\mathbf{b}, \mathbf{a} + 2\mathbf{b} \rangle &= 2 \langle \mathbf{a}, \mathbf{a} \rangle + \langle \mathbf{a}, \mathbf{b} \rangle - 6 \langle \mathbf{b}, \mathbf{b} \rangle = \\ &= 2 \cdot 46 + 7 - 6 \cdot 17 = -3, \end{aligned}$$

where $\langle \mathbf{a}, \mathbf{a} \rangle = 3^2 + (-6)^2 + 1^2 = 46$,
$\langle \mathbf{a}, \mathbf{b} \rangle = 3 \cdot 1 + (-6) \cdot 0 + 1 \cdot 4 = 7$ and
$\langle \mathbf{b}, \mathbf{b} \rangle = 1^2 + 0^2 + 4^2 = 17$.

Using the concept of inner product the *norm* of a vector can be defined:

Definition (Norm)

For a vector in $\mathbb{R}^n$ $\|\mathbf{x}\| := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}$.

The norm of a vector is obviously the generalization of the absolute value of real numbers, or the length of vectors in the 2-dimensional Cartesian coordinate systems.

It is easy to verify that the following two important properties of function norm hold true.

1. $\|\mathbf{x}\| \geq 0$ and $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = \mathbf{0}$,
2. $\|\lambda \mathbf{x}\| = |\lambda| \cdot \|\mathbf{x}\|$.

One can see the first two properties. For the third one we need the so-called *Cauchy–Schwarz inequality*.

Theorem (Cauchy–Schwarz inequality)

For any pair $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|.$$

We have for any real number λ

$$\langle \mathbf{x} - \lambda\mathbf{y}, \mathbf{x} - \lambda\mathbf{y} \rangle = \|\mathbf{x}\|^2 - 2\lambda \langle \mathbf{x}, \mathbf{y} \rangle + \lambda^2 \|\mathbf{y}\|^2 \geq 0,$$

This is a quadratic polynomial in λ which is non-negative everywhere. Therefore its discriminant is not positive:

$$D = 4(\langle \mathbf{x}, \mathbf{y} \rangle)^2 - 4\|\mathbf{x}\|^2 \|\mathbf{y}\|^2 \leq 0.$$

It was to be proved.

Triangle inequality

We are ready to prove the triangle inequality and summarize the three most important properties of the norm.

Theorem (Properties of norm)

1. $\|\mathbf{x}\| \geq 0$ *and* $\|\mathbf{x}\| = 0$ *iff* $\mathbf{x} = \mathbf{0}$,
2. $\|\lambda \mathbf{x}\| = |\lambda| \cdot \|\mathbf{x}\|$,
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

The first two properties hold true as we saw it earlier.
Now we have

$$\begin{aligned}\|\mathbf{x}+\mathbf{y}\|^2 = \langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y} \rangle = \|\mathbf{x}\|^2 + 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2 \leq \\ \leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\| \cdot \|\mathbf{y}\| + \|\mathbf{y}\|^2 = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2 .\end{aligned}$$

Using the concept of the norm of vectors a distance function can be introduced.

Definition

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n : d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|.$$

From the properties of the norm of vectors the expected properties of the distance function immediately follow:

Properties of the distance function

1. $d(\mathbf{x}, \mathbf{y}) \geq 0$ and $d(\mathbf{x}, \mathbf{y}) = 0$, iff $\mathbf{x} = \mathbf{y}$,
2. $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$,
3. $d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$.

The Cauchy–Schwarz inequality implies that

$$-1 \leq \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|} \leq 1.$$

Therefore there is a unique angle $\varphi \in [0, \pi]$, such that

$$\cos \varphi = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|}.$$

Definition (Angle)

The number $\varphi \in [0, \pi]$ defined above is the *angle* enclosed by the vectors $\mathbf{x}$ and $\mathbf{y}$.

Introducing the inner product in $\mathbb{R}^n$ it became possible to define distance of vectors and measuring the angle of vectors. Therefore the Euclidean geometry of the plane or space can be generalized. Therefore the vector space $\mathbb{R}^n$ with the inner product is said to be *Euclidean n-space.*

Example

Find the area of triangle with vertices $\mathbf{a} = (1, 1 - 1, 1)$, $\mathbf{b} = (2, 1, -1, 0)$ and $\mathbf{c} = (1, 2, 1, 1)$.

The length of the sides in the triangle are:
$d(\mathbf{a}, \mathbf{b}) = \|\mathbf{b} - \mathbf{a}\| = \|(1, 0, 0, o - 1)\| = \sqrt{2}$, $d(\mathbf{a}, \mathbf{c}) = \|\mathbf{c} - \mathbf{a}\| = \|(0, 1, 2, 0)\| = \sqrt{5}$, $d(\mathbf{b}, \mathbf{c}) = \|\mathbf{c} - \mathbf{b}\| = \|(1, 0, 0, 1)\| = \sqrt{7}$. The angle at the vertex $\mathbf{a}$ is 90°, because $\langle \mathbf{b} - \mathbf{a}, \mathbf{c} - \mathbf{a} \rangle = 0$, therefore the area of the triangle is $\frac{\sqrt{10}}{2}$.

Definition (Orthonormal system of vectors)

Two vectors $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$ are said to be *orthogonal* if their inner product is zero.

The set of vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ is an orthonormal set, if

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \begin{cases} 0 & \text{if} \quad i \neq j, \\ 1 & \text{if} \quad i = j. \end{cases}$$

Theorem

An orthonormal set is linearly independent.

Definition (Orthonormal basis)

If any vector $\mathbf{v} \in \mathbb{R}^n$ is the linear combination of the vectors of the orthonormal set $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, then $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ is called *orthonormal basis*, or Cartesian coordinate system.

The set $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, is a Cartesian coordinate system of the Euclidean n-space.

Concept of Matrices

In our everyday life we often meet tables of numbers containing statistical data, coefficients of system of linear equations and so on. In linear algebra these tables of numbers are referred to as matrices.

Definition

A rectangular shaped array of numbers is called matrix, that is, a matrix is

$$\mathbf{A} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \dots & \alpha_{mn} \end{bmatrix},$$

where α_{ij} $(i = 1, \dots, m;\ j = 1, \dots, n)$ are numbers. If the matrix has m rows and n columns, then it is said to be of type $m \times n$. If $m = n$ then we call it square matrix of *order* m. Here α_{ij} is the entry in the intersection of the ith row and jth column.

If the entries of the matrix are real numbers, then we say that $\mathbf{A}$ is a real matrix, and the set of all real matrices of type $m \times n$ will be denoted by $\mathbb{R}^{m \times n}$.

For short we often refer to a matrix $\mathbf{A}$ by its general entry and we denote it as $[\alpha_{ij}]$.
If a matrix has only one column, then we call it column matrix or more often column vector and similarly if the matrix consists of only one row, then we call it row matrix or row vector.
In some cases we partition the matrix into sub-matrices

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \dots & \mathbf{A}_{1k} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \dots & \mathbf{A}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{\ell 1} & \mathbf{A}_{\ell 2} & \dots & \mathbf{A}_{\ell k} \end{bmatrix}$$

In particular, we often use the partition, where the sub-matrices are the columns or rows of the matrix

$$\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n] = \begin{bmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_m \end{bmatrix}$$

where $\mathbf{a}_i$ denotes the ith column and $\mathbf{a}'_j$ the jth row of $\mathbf{A}$.

Matrix operations

Operations on $\mathbb{R}^{m\times n}$

The sum of two matrices of the same type $\mathbf{A} = [\alpha_{ij}]$ and $\mathbf{B} = [\beta_{ij}]$ is the matrix

$$\mathbf{A} + \mathbf{B} = [\alpha_{ij} + \beta_{ij}].$$

If $\lambda \in \mathbb{R}$ and $\mathbf{A} = [\alpha_{ij}] \in \mathbb{R}^{m\times n}$ then we define the multiplication of a matrix by scalar $\lambda\mathbf{A}$ as follows:

$$\lambda\mathbf{A} = [\lambda\alpha_{ij}].$$

Combining the addition and multiplication by scalars we can form the linear combination of matrices of the same type: $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ and $\mathbf{A}_1, \dots, \mathbf{A}_k \in \mathbb{R}^{m\times n}$, then their linear combination is

$$\lambda_1\mathbf{A}_1 + \cdots + \lambda_k\mathbf{A}_k \in \mathbb{R}^{m\times n}.$$

Example

If $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 6 & 5 & 4 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} -1 & 0 & 1 \\ 2 & -2 & 3 \end{bmatrix}$ then the linear combination $2\mathbf{A} - 3\mathbf{B}$ is

$$
\begin{aligned}
2\mathbf{A} - 3\mathbf{B} &= \underbrace{\begin{bmatrix} 2 & 4 & 6 \\ 12 & 10 & 8 \end{bmatrix}}_{2\mathbf{A}} - \underbrace{\begin{bmatrix} -3 & 0 & 3 \\ 6 & -6 & 9 \end{bmatrix}}_{3\mathbf{B}} = \\
&= \begin{bmatrix} 2-(-3) & 4-0 & 6-3 \\ 12-6 & 10-(-6) & 8-9 \end{bmatrix} = \\
&= \begin{bmatrix} 5 & 4 & 3 \\ 6 & 16 & -1 \end{bmatrix}.
\end{aligned}
$$

Similarly to the operations of vectors, the addition of matrices is also led back to the addition of numbers, and the multiplication of matrices by scalars is led back to the multiplication of numbers, therefore the following properties are obviously hold true. If $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are arbitrary matrices in $\mathbb{R}^{m \times n}$ and $\alpha, \beta \in \mathbb{R}$, then

1. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$,
2. $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$,
3. $\exists \mathbf{O} : \forall \mathbf{A} : \mathbf{O} + \mathbf{A} = \mathbf{A}$,
4. $\forall \mathbf{A} : \exists (-\mathbf{A}) : \mathbf{A} + (-\mathbf{A}) = \mathbf{O}$.

Vector space of matrices

The addition of matrices is commutative, associative, there exists zero matrix (each entry is 0), and each matrix has an opposite (additive inverse).

1. $\alpha(\mathbf{A}+\mathbf{B}) = \alpha\mathbf{A} + \alpha\mathbf{B}$,
2. $(\alpha+\beta)\mathbf{A} = \alpha\mathbf{A} + \beta\mathbf{A}$,
3. $(\alpha\beta)\mathbf{A} = \alpha(\beta\mathbf{A})$,
4. $1\mathbf{A} = \mathbf{A}$.

The set $\mathbb{R}^{m\times n}$ of matrices with the addition and scalar multiplication is a real vector space.

Multiplication of matrices

While the addition of matrices was defined for matrices having the same type, the multiplication of two matrices is defined only if the number of columns in the left side matrix equals the number of rows in the right side one.

Definition (Product of matrices)

Let $\mathbf{A} = [\alpha_{ij}]$ be an $m \times n$ matrix and $\mathbf{B} = [\beta_{jk}]$ be an $n \times p$ matrix, then their product $\mathbf{AB} = \mathbf{C} = [\gamma_{ij}]$ is the $m \times p$ matrix for which

$$\gamma_{ij} = \sum_{k=1}^{n} \alpha_{ik}\beta_{kj} = \alpha_{i1}\beta_{1j} + \alpha_{i2}\beta_{2j} + \cdots + \alpha_{in}\beta_{nj}.$$

Example

If $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 6 & 5 & 4 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} -1 & 0 \\ 2 & -2 \\ 0 & 1 \end{bmatrix}$ then their product $\mathbf{AB}$ is

$$\begin{bmatrix} 1 \cdot (-1) + 2 \cdot 2 + 3 \cdot 0 & 1 \cdot 0 + 2 \cdot (-2) + 3 \cdot 1 \\ 6 \cdot (-1) + 5 \cdot 2 + 4 \cdot 0 & 6 \cdot 0 + 5 \cdot (-2) + 4 \cdot 1 \end{bmatrix} = \begin{bmatrix} 3 & -11 \\ 4 & -6 \end{bmatrix}.$$

Matrix multiplied by a vector from right

Ax

Multiplying the matrix of type $m \times n$

$$\mathbf{A} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \dots & \alpha_{mn} \end{bmatrix} \quad \text{by the vector} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

the product is the linear combination of the columns of $\mathbf{A}$, where the scalar coefficients are the components of $\mathbf{x}$.

Proof

$$\mathbf{Ax} = \begin{bmatrix} x_1\alpha_{11} + x_2\alpha_{12} + \cdots + x_n\alpha_{1n} \\ x_1\alpha_{21} + x_2\alpha_{22} + \cdots + x_n\alpha_{2n} \\ \vdots \\ x_1\alpha_{m1} + x_2\alpha_{m2} + \cdots + x_n\alpha_{mn} \end{bmatrix} =$$

$$= x_1 \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \\ \vdots \\ \alpha_{m1} \end{bmatrix} + x_2 \begin{bmatrix} \alpha_{12} \\ \alpha_{22} \\ \vdots \\ \alpha_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} \alpha_{1n} \\ \alpha_{2n} \\ \vdots \\ \alpha_{mn} \end{bmatrix},$$

Matrix multiplied by a vector from right

xB

The product of a row vector and a matrix is a row vector, which is a linear combination of the rows of the matrix.

If $\mathbf{A} = \begin{bmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_m \end{bmatrix}$, where $\mathbf{a}'_i$ denotes the ith row of $\mathbf{A}$ and $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_p]$, where $\mathbf{b}_j$ is the jth column of $\mathbf{B}$, then their product is

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a}'_1\mathbf{b}_1 & \mathbf{a}'_1\mathbf{b}_2 & \dots & \mathbf{a}'_1\mathbf{b}_p \\ \mathbf{a}'_2\mathbf{b}_1 & \mathbf{a}'_2\mathbf{b}_2 & \dots & \mathbf{a}'_2\mathbf{b}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}'_m\mathbf{b}_1 & \mathbf{a}'_m\mathbf{b}_2 & \dots & \mathbf{a}'_m\mathbf{b}_p \end{bmatrix} = [\mathbf{A}\mathbf{b}_1, \mathbf{A}\mathbf{b}_2, \dots, \mathbf{A}\mathbf{b}_p] =$$

$$= \begin{bmatrix} \mathbf{a}'_1\mathbf{B} \\ \mathbf{a}'_2\mathbf{B} \\ \vdots \\ \mathbf{a}'_m\mathbf{B} \end{bmatrix}$$

The ijth entry $\mathbf{a}'_i \mathbf{b}_j$ of the product is the product of the row matrix $\mathbf{a}'_i$ and column matrix $\mathbf{b}_j$, which is a scalar like the inner product of two vectors.

Theorem

Each column of the matrix product $\mathbf{AB}$ *is the linear combination of the columns of* $\mathbf{A}$ *and each row of* $\mathbf{AB}$ *is the linear combination of the rows of* $\mathbf{B}$.

Theorem (Properties of matrix multiplication)

1. *noncommutative,*
2. *associative,*
3. *distributive with respect to the addition.*

Powers of matrix

Power of square matrix

Using the multiplication of matrices we can define the powers of square matrices (number of rows and columns are equal) with non-negative integer exponents as follows: if $\mathbf{A} \in \mathbb{R}^{n\times n}$ then

$$\mathbf{A}^0 := \mathbf{E}_n \text{ and } \forall k \in \mathbb{N} : \mathbf{A}^{k+1} := \mathbf{A}^k \cdot \mathbf{A}.$$

where $\mathbf{E}_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$ is the $n \times n$ identity matrix. Notice that for an arbitrary matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$: $\mathbf{A}\mathbf{E}_n = \mathbf{E}_m\mathbf{A} = \mathbf{A}$, that is, the identity matrix plays the same role among the matrices as the 1 among the numbers.

- Let
$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$
Calculate $A^2, A^3, A^4, \cdots$ until you detect a pattern. Write a general formula for A^n.

Problems

- Let
$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$
Calculate $A^2, A^3, A^4, \cdots$ until you detect a pattern. Write a general formula for A^n.
- Let
$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$
Calculate $A^2, A^3, A^4, \cdots$ until you detect a pattern. Write a general formula for A^n.

The rank

Transpose of matrices

Definition

The transpose of a matrix $\mathbf{A} = [\alpha_{ij}] \in \mathbb{R}^{m \times n}$ is the matrix $\mathbf{A}^\top$ of type $n \times m$ is obtained by exchanging $\mathbf{A}'$s columns and rows.

Properties:

1. $(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$,
2. $(\lambda \mathbf{A})^\top = \lambda \mathbf{A}^\top$,
3. $(\mathbf{A}\mathbf{B})^\top = \mathbf{B}^\top \mathbf{A}^\top$.

Definition

1. The rank of a system of vectors is the dimension of the span of this system.
2. The column space of a matrix is span of the column system of the matrix. Similarly, the row space of a matrix is the span of row system of the matrix.
3. The column rank of a matrix is the dimension of the column space. Similarly, the row rank of a matrix is the dimension of the row space.

Our goal is to understand that for any matrices the two concepts of rank, the column rank and the row rank, are the same!