# Some basic definitions and facts from linear algebra[1]

**A matrix** $m$-by-$n$ is a table of numbers, having $m$ rows and $n$ columns. The **element**, or **entry**, $a_{ij}$ of a matrix $A$ is the number in row $i$ and column $j$, counting from top-left. Sometimes, especially for particular values of $i$ and $j$, we write $a_{i,j}$ instead of $a_{ij}$. For example, $a_{2,1}$ is the same as $a_{21}$. It is a representation of a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$.

**A transpose** $A^\top$ of a matrix $A = (a_{ij})$ is the matrix $A^\top = (a_{ji})$.

**A vector** is a matrix with only one column; the size of the vector is the number of rows.

**The product** of an $m$-by-$n$ matrix $A = (a_{ij})$ with an $n$-by-$k$ matrix $B = (b_{j\ell})$ is an $m$-by-$k$ matrix with the element in the row $i$ and column $\ell$ equal to

$$\sum_{j=1}^{n} a_{ij} b_{j\ell}$$

**A square matrix** has equal number or rows and columns.

**The diagonal elements** of a square matrix $A = (a_{ij})$ are $a_{ii}$. The element $a_{ij}$ is said to be above the diagonal if $i < j$ (e.g. $a_{1,2}$) and below the diagonal if $i > j$ (e.g. $a_{2,1}$).

**The trace** of a square matrix is the sum of the diagonal elements:

$$\mathrm{Tr}(A) = a_{11} + a_{22} + \cdots .$$

**The determinant** of a square $n$-by-$n$ matrix $A = (a_{ij})$ is the number

$$\det(A) = |A| = \sum_{\sigma} (-1)^{|\sigma|} a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}$$

where summation is carried out over all $n!$ permutations of the set $(1, 2, \ldots, n)$, and $|\sigma|$ is the number of times a bigger number comes in front of the smaller number in the sequence $(\sigma(1), \sigma(2), \ldots, \sigma(n))$.

**The identity matrix** $I$ is a square matrix with all diagonal elements equal to 1 and all other elements equal to zero. Writing $I_n$ signifies that $I$ has $n$ rows and $n$ columns.

**A symmetric matrix** $A = (a_{ij})$ is a square matrix with $a_{ij} = a_{ji}$.

**A skew-symmetric matrix** $A = (a_{ij})$ is a square matrix with $a_{ij} = -a_{ji}$. In particular, the diagonal elements of a skew-symmetric matrix are equal to zero.

**A singular matrix** is a square matrix with zero determinant.

**A diagonal matrix** is a square matrix whose non-zero elements are only on the diagonal: $a_{ij} = 0$ if $i \neq j$.

**An upper triangular matrix** has $a_{ij} = 0$ for $i > j$ (zeroes below the diagonal).

**A lower triangular matrix** has $a_{ij} = 0$ for $i < j$ (zeroes above the diagonal).

**Elementary row operations** on the matrix with rows $r_1, \ldots, r_n$ are

(1) exchanging the rows: $r_i \leftrightarrow r_j$;
(2) multiplying a row by a non-zero number $c$: $r_i \leftarrow c r_i$;
(3) adding to a row a multiple of another row: $r_i \leftarrow r_i + c r_j$.

**Gaussian elimination** is a special procedure for solving a linear system of equations using elementary row operations.

**Linear** (or **vector**) space $V$ over the real or complex numbers is a collection of objects that can be added and multiplied by a number so that all the "obvious" properties hold (but must be explicitly required): for $u, v, w \in V$ and numbers $a, b$,

(1) commutativity of vector addition: $u + v = v + u$;
(2) associativity of vector addition: $(u + v) + w = u + (v + w)$;
(3) existence of zero vector $\theta \in V$ such that $u + \theta = u$ for all $u \in V$;
(4) existence of the additive inverse: for every $u \in V$, there is a (unique) $-u \in V$ such that $u + (-u) = \theta$;
(5) distributivity laws: $a(u + v) = au + av$, $(a + b)u = au + bu$, $(ab)u = a(bu)$;
(6) one more: $1u = u$.

An immediate exercise after this definition is to prove some other "obvious" properties, such as $0u = \theta$ and $(-1)u = -u$.

Main examples of vector spaces are (a) $\mathbb{R}^{mn}$: matrices of the given size ($m$ rows, $n$ columns), of which the (proper) vectors $\mathbb{R}^m$ are an example, and (b) $\mathcal{C}^n([a, b])$: continuous functions defined on a given interval $[a, b]$ and having $n \geq 0$ continuous derivatives.

[1] Sergey Lototsky, USC, version of December 7, 2023

**A linear combination** of elements $u_1, \ldots, u_n$ of a vector space is an element of the form $a_1u_1 + \ldots + a_nu_n$, where $a_1, \ldots, a_n$ are some numbers.

**A linear span** of elements $u_1, \ldots, u_n$ of a vector space is the collection of all possible linear combinations of these elements, that is, the collection of $a_1u_1 + \ldots + a_nu_n$, for all possible values of the numbers $a_1, \ldots, a_n$.

The elements $u_1, \ldots, u_n$ of a vector space are called **linearly dependent** if it is possible to get $a_1u_1 + \ldots + a_nu_n = \theta$ with not all numbers $a_k$ equal to zero. Otherwise, the elements are called **linearly independent**.

The **dimension** of a vector space is the largest number of linearly independent elements in that space. In particular, the space $\mathbb{R}^n$ is exactly $n$-dimensional. As a result, any four vectors in $\mathbb{R}^3$ are linearly dependent.

**The rank** of a matrix is the maximal number of linearly independent rows (or columns). The rank of a rectangular $m$-by-$n$ matrix cannot exceed the *smaller* of the two numbers $n$ and $m$. A non-singular square matrix of size $n$ has the **full rank** $n$. If $u$ and $v$ are two non-zero column vectors of size $n$, then the $n$-by-$n$ matrix $uv^\top$ always has rank equal to 1. **The kernel** (or null space) of a matrix $A$ is the collection of vectors $u$ such that $Au = 0$; it is a linear space. For every rectangular $m$-by-$n$ matrix, the number $n$ of *columns* is equal to the rank plus the dimension of the kernel: this is the **rank-nullity theorem**. In the operator language, the dimension $n$ of the domain of a linear mapping is the sum of the dimensions of the range and of the kernel.

The **characteristic polynomial** $P_A(\lambda)$ of an $n$-by-$n$ square matrix $A$ is

$$P_A(\lambda) = \det(A - \lambda I),$$

where $I$ is the identity matrix of the same size as $A$. The degree of this polynomial is $n$ An **eigenvalue** of the matrix $A$ is a root of the characteristic polynomial of $A$. Complex eigenvalues are allowed, and, by the fundamental theorem of algebra, we have

$$P_A(\lambda) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_k)^{n_k},$$

where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of $A$. The corresponding number $n_i$ is called the **algebraic multiplicity** of the eigenvalue $\lambda_i$. Note that $n_1 + n_2 + \ldots + n_k = n$.

*The sum of all eigenvalues, counting multiplicity, is equal to the trace of the matrix. The product of all eigenvalues, counting multiplicity, is equal to the determinant of the matrix.* In particular, a matrix is singular if and only if it has a zero eigenvalue.

**The eigenvector** $v_i$ of the matrix $A$, corresponding to the eigenvalue $\lambda_i$, is a non-zero solution of the equation

$$(A - \lambda_i)v_i = \theta \ (\theta \ \text{ is the zero vector}),$$

or, equivalently,

$$Av_i = \lambda_i v_i.$$

The **geometric multiplicity** of the eigenvalue $\lambda_i$ is the number of linearly independent eigenvectors corresponding to it. The geometric multiplicity is not bigger than the algebraic multiplicity. If the geometric multiplicity is strictly less than the algebraic multiplicity, then the eigenvalue is called **defective**. A matrix is called defective if it has at least one defective eigenvalue.

**A generalized eigenvector** $u$ of the matrix $A$, corresponding to the eigenvalue $\lambda$ with algebraic multiplicity $m > 1$, is a non-zero solution of the equation

$$(A - \lambda I)^m u = 0.$$

**Theorem.** (a) If $\lambda$ is an eigenvalue of the $n$-by-$n$ matrix $A$ and has algebraic multiplicity $m$, then the rank of the matrix $(A - \lambda I)^m$ is $n - m$ and the equation

$$(A - \lambda I)^m v = 0.$$

has exactly $m$ *linearly independent* solutions $u_1, \ldots, u_m$.

(b) If $A = A^\top$, then all eigenvalues $\lambda_k$ of $A$ are real and non-defective, and the corresponding eigenvectors $v_k$ can be arranged into an orthonormal basis in $\mathbb{R}^n$ so that $A = \sum_k \lambda_k v_k v_k^\top = V \Lambda V^\top$ with a diagonal matrix $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and *orthogonal* matrix $V = [v_1 \ldots v_n]$.