Lecture 1: Generalities on vectors and matrices

Francesco Preta

August 18, 2020

Vectors and Matrices

Vectors in $\mathbb{R}^n$ and their linear combinations

In this class we will mostly deal with operations in $\mathbb{R}^n$. For our purposes, vectors will be defined as ordered lists of real numbers which can assume different interpretations depending on the context. The notation we will use will be either the one of row vectors, (horizontal lists of real numbers) in square brackets such as

$$\begin{bmatrix} 1 & 2 & \pi \end{bmatrix}$$

or column vectors such as

$$\begin{bmatrix} 1 \\ 2 \\ \pi \end{bmatrix}$$

Sometimes we will use lists like $(1, 2, \pi)$ to indicate vectors. In this specific case, we will treat them as coordinates of a point in $\mathbb{R}^n$.

Different sources identify vectors with different notations. Sometimes they will be identified by a letter with an arrow on top of it ($\vec{v}$). In our case, we will use boldface ($\mathbf{v}$) to identify vectors and uppercase letter to identify matrices. Scalars will generally be identified by greek letters.

Each number in the vector is called an *element* of the vector and the number of elements is called *size* or *dimension*.

A variety of objects can be represented through vectors. For instance a 3-dimensional vector can represent a position in space, in which each coordinate represents the distance from a fixed origin on a different axis.

Alternatively, a vector can represent a set of features characterizing a data point. The position of each component will determine which feature it refers to.

There two fundamental elementary operations that can be computed on vectors: addition and multiplication by a scalar.

First, the *sum*, or *addition* of two vectors of the same dimension n is a n-dimensional vector whose components are the sums of the corresponding components of the original vectors. That is, let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, then

$$\mathbf{v} + \mathbf{w} = \begin{bmatrix} v_1 \\ v_2 \\ ... \\ v_n \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ ... \\ w_n \end{bmatrix} = \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \\ ... \\ v_n + w_n \end{bmatrix}$$

or, in coordinate notation:

$$(\mathbf{v} + \mathbf{w})_i = \mathbf{v}_i + \mathbf{w}_i \quad \text{for} \quad i = 1, ..., n$$

For example

$$\begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} + \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \\ 6 \end{bmatrix}$$

Vector addition follows the following properties, for every $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$:

1. **Associativity**: $(\mathbf{v} + \mathbf{w}) + \mathbf{u} = \mathbf{v} + (\mathbf{w} + \mathbf{u})$.

2. **Commutativity**: $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$.

3. **Existence of Identity element**: if $\mathbf{0}$ is a n-dimensional vector of zeroes, $\mathbf{0} + \mathbf{v} = \mathbf{v}$.

4. **Existence of inverse**: $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$ where $-\mathbf{v}$ is the vector such that $(-\mathbf{v})_i = -\mathbf{v}_i$.

In dimension 2 or 3, the geometric interpretation of vector addition can be visualized by considering the plane containing the two vectors. These can be interpreted as two arrows from the origin to their respective coordinates. The vector corresponding to their sum, will be the diagonal of the parallelogram spanned by these two vectors which intersects the origin, like in Figure 1

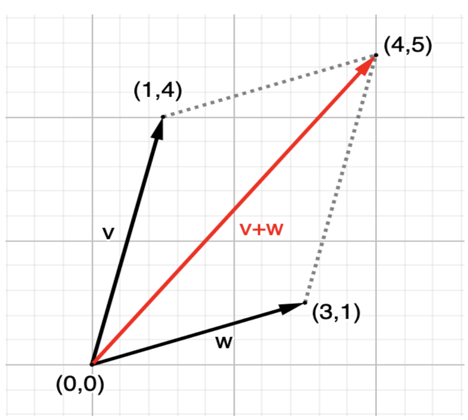

Figure 1: Parallelogram rule for the sum of vectors $\mathbf{v} = \begin{bmatrix} 1 & 4 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 3 & 1 \end{bmatrix}$

Another fundamental elementary operation is *scalar multiplication*. We call a *scalar* any real number $\alpha \in \mathbb{R}$. For any $\mathbf{v} \in \mathbb{R}^n$, $\alpha\mathbf{v}$ is the vector obtained by multiplying each component of $\mathbf{v}$ by α, that is

$$\alpha\mathbf{v} = \begin{bmatrix} \alpha v_1 \\ \alpha v_2 \\ \alpha v_3 \end{bmatrix}$$

For example, for $\alpha = 2$ and $\mathbf{v} = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$ we get

$$\alpha\mathbf{v} = 2\begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \\ 12 \end{bmatrix}$$

The following are properties of scalar multiplication, for every $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$:

1. **Compatibility with field multiplication**: $(\alpha\beta)\mathbf{v} = \alpha(\beta\mathbf{v})$.
2. **Existence of identity element**: $1\mathbf{v} = \mathbf{v}$.
3. **Distributivity with respect to vector addition**: $\alpha(\mathbf{v}+\mathbf{w}) = \alpha\mathbf{v}+\alpha\mathbf{w}$.
4. **Distributivity with respect to real addition**: $(\alpha+\beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$.

Notice that scalar multiplication has the same properties if we choose scalars to be complex rather than real numbers.

The geometric interpretation of scalar multiplication is the following: if a vector $\mathbf{v}$ is identified as an arrow from the origin to its coordinates, any scalar multiple of $\mathbf{v}$ will also lie on the same line, with an opposite direction if the scalar is negative.

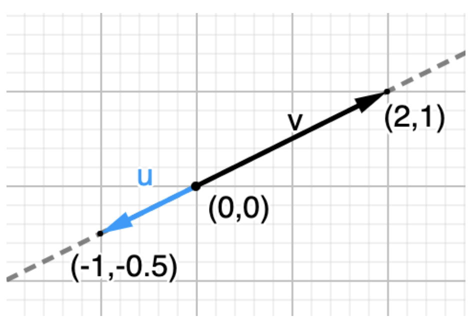

Figure 2: Geometric representation of scalar multiplication for $\mathbf{v} = \begin{bmatrix} 1 & 2 \end{bmatrix}$ and $\mathbf{u} = -0.5\mathbf{v}$.

Now that we have defined vector addition and scalar multiplication, we can combine them to get the notion of *linear combination*:

Definition 1. Let $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ be a set of vectors in $\mathbb{R}^n$. A linear combination of $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ is any vector of the form

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + ... + \alpha_k\mathbf{v}_k$$

for some choice of $\alpha_1, ..., \alpha_k \in \mathbb{R}$.

For instance, let $\mathbf{v}_1 = \begin{bmatrix}1 & 2 & 0\end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix}-1 & 1 & 0\end{bmatrix}$. Then the vector

$$\mathbf{w} = 3\mathbf{v}_1 - 2\mathbf{v}_2 = \begin{bmatrix}5 & 4 & 0\end{bmatrix}$$

is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. It is generally an interesting question whether a given vector $\mathbf{u}$ is a linear combination of an existing set of vectors. In our example, a vector like $\mathbf{u} = \begin{bmatrix}0 & 0 & 1\end{bmatrix}$ cannot be a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ since any linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ will have third coordinate equal to 0. To solve this question in its full generality, we will need to develop tools to solve systems of linear equations, which will be done in future lectures.

The notion of linear combination gives rise to another important notion:

Definition 2. Let $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ be a set of vectors in $\mathbb{R}^n$. The *span* of $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ (also called *linear subspace spanned by* $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$) is the set

$$\mathrm{span}(\mathbf{v}_1, ..., \mathbf{v}_k) = \left\{\mathbf{v} \in \mathbb{R}^n \mid \mathbf{v} = \sum_{i=1}^{k} \alpha_i\mathbf{v}_i \text{ for some } \alpha_1, ..., \alpha_k \in \mathbb{R}\right\}$$

The previous question of whether a given vector $\mathbf{u}$ is a linear combination of our vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ can then be rephrased by asking if $\mathbf{u}$ is in the span of $\mathbf{v}_1$ and $\mathbf{v}_2$.

In $\mathbb{R}^n$ we can define a set $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ called the *canonical basis*. Each $\mathbf{e}_i$ is a vector of all 0s, except for the i-th coordinate which is equal to 1. We have that $\mathbb{R}^n = \mathrm{span}(\mathbf{e}_1, ..., \mathbf{e}_n)$ since for every $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{x} = \begin{bmatrix}x_1 \\ x_2 \\ ... \\ x_n\end{bmatrix} = x_1\begin{bmatrix}1 \\ 0 \\ ... \\ 0\end{bmatrix} + x_2\begin{bmatrix}0 \\ 1 \\ ... \\ 0\end{bmatrix} + ... + x_n\begin{bmatrix}0 \\ 0 \\ ... \\ 1\end{bmatrix} = \sum_{i=1}^{n} x_i\mathbf{e}_i$$

so every element of $\mathbb{R}^n$ can be written as a linear combination of the elements of the canonical basis.

Abstract vector spaces

Up until this point, we have defined a vector as an element of $\mathbb{R}^n$ and we have listed its property with respect to two important operations: vector addition and scalar multiplication. The properties of this type of space can be generalized through the following definition:

Definition 3. A *vector space* V over $\mathbb{R}$ is a set, along with operations $(+, \cdot)$, such that for every $\mathbf{v}, \mathbf{w}, \mathbf{u}$ in V and $\alpha, \beta \in \mathbb{R}$, the following holds:

1. **Associativity of +**: $(\mathbf{v}+\mathbf{w})+\mathbf{u}=\mathbf{v}+(\mathbf{w}+\mathbf{u})$.
2. **Commutativity of +**: $\mathbf{v}+\mathbf{w}=\mathbf{w}+\mathbf{v}$.
3. **Existence of Identity element for +**: there exists an element $\mathbf{0}$ such that $\mathbf{0}+\mathbf{v}=\mathbf{v}$.
4. **Existence of inverse for +**: for every $\mathbf{v}\in V$ there exists an element $\mathbf{v}'$ (called the inverse of $\mathbf{v}$) such that $\mathbf{v}+(\mathbf{v}')=\mathbf{0}$.
5. **Compatibility of $\cdot$ with field multiplication**: $(\alpha\beta)\cdot\mathbf{v}=\alpha(\beta\cdot\mathbf{v})$.
6. **Existence of identity element for $\cdot$**: $1\cdot\mathbf{v}=\mathbf{v}$.
7. **Distributivity of $\cdot$ with respect to +**: $\alpha\cdot(\mathbf{v}+\mathbf{w})=\alpha\cdot\mathbf{v}+\alpha\cdot\mathbf{w}$.
8. **Distributivity of $\cdot$ with respect to real addition**: $(\alpha+\beta)\cdot\mathbf{v}=\alpha\cdot\mathbf{v}+\beta\cdot\mathbf{v}$.

Moreover, we require that for every $\mathbf{v},\mathbf{w}\in V$ and $\alpha,\beta\in\mathbb{R}$, $\alpha\mathbf{v}+\beta\mathbf{w}\in V$.

In practice, we define a *vector space* to be a set V with two operations that resemble vector addition and scalar multiplication in $\mathbb{R}^n$ and such that V is closed under linear combinations.

Example 1:

$$\mathcal{C}[0,1]=\{f:[0,1]\to\mathbb{R}\,|\ \text{f is continuous }\}$$

In this case, scalar multiplication $\cdot$ is just pointwise multiplication of the function by a scalar: $(\alpha\cdot f)(x)=\alpha f(x)$, while addition + is point-wise addition between functions $(f_1+f_2)(x)=f_1(x)+f_2(x)$. You can check that these operations satisfy the axioms we just listed in the definition. Moreover, if $f_1,f_2\in\mathcal{C}([0,1])$, then for every $\alpha,\beta\in\mathbb{R}$, $\alpha f_1+\beta f_2$ is also a continuous function from $[0,1]$ to $\mathbb{R}$. Therefore $\mathcal{C}[0,1]$ is a vector space.

Example 2:
For a given set of vectors in $\mathbb{R}^n$, $\{\mathbf{v}_1,...,\mathbf{v}_k\}$, $\text{span}(\mathbf{v}_1,...,\mathbf{v}_k)$ is a vector space. Since operations are inherited from $\mathbb{R}^n$, they satisfy all the axioms of the definition. Moreover, $\text{span}(\mathbf{v}_1,...,\mathbf{v}_k)$ is closed under linear combinations, that is, suppose $\mathbf{u},\mathbf{w}\in\text{span}(\mathbf{v}_1,...,\mathbf{v}_k)$. We want to prove that for every $\gamma,\delta\in\mathbb{R}$, $\gamma\mathbf{u}+\delta\mathbf{w}\in\text{span}(\mathbf{v}_1,...,\mathbf{v}_k)$. That means that we need to find a set of real numbers $(\eta_1,...,\eta_k)$ such that

$$\gamma\mathbf{u}+\delta\mathbf{w}=\sum_{i=1}^{k}\eta_i\mathbf{v}_i$$

Since $\mathbf{u}$ and $\mathbf{w}$ are elements of $\text{span}(\mathbf{v}_1,...,\mathbf{v}_k)$, there exist $(\alpha_1,...,\alpha_k)$ and $(\beta_1,...,\beta_k)$ such that

$$\mathbf{u}=\sum_{i=1}^{k}\alpha_i\mathbf{v}_i,\qquad \mathbf{w}=\sum_{i=1}^{k}\beta_i\mathbf{v}_i$$

But then by using properties of vector addition and scalar multiplication,

$$\gamma \mathbf{u} + \delta \mathbf{w} = \gamma \sum_{i=1}^{k} \alpha_i \mathbf{v}_i + \delta \sum_{i=1}^{k} \beta_i \mathbf{v}_i = \sum_{i=1}^{k} (\gamma \alpha_i + \delta \beta_i) \mathbf{v}_i$$

so that the result is proved for $\eta_i = \gamma \alpha_i + \delta \beta_i$.

Scalar product, norm and cosine

We define now a few more operations between vectors in $\mathbb{R}^n$ that we will encounter in our class. The first is given by the *scalar product* (also called dot product, or inner product). For any pair of vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$ their scalar product is defined as

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i=1}^{n} v_i w_i$$

For instance,

$$\langle \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} \rangle = 2 \cdot 3 + 4 \cdot 2 + 6 \cdot 0 = 14$$

Notice that although $\mathbf{v}$ and $\mathbf{w}$ are vectors, their scalar product is a scalar. For every $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$, the scalar product has the following properties:

1. **Linearity in the first component**:$\langle \alpha \mathbf{v} + \beta \mathbf{w}, \mathbf{u} \rangle = \alpha \langle \mathbf{v}, \mathbf{u} \rangle + \beta \langle \mathbf{w}, \mathbf{u} \rangle$.
2. **Symmetry**:$\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$.
3. **Positive Definite**:$\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ and it is equal to 0 if and only if $\mathbf{v} = \mathbf{0}$.

Related to the scalar product is the notion of *Euclidean norm* of a vector in $\mathbb{R}^n$, defined as

$$||\mathbf{v}|| = \sqrt{\sum_{i=1}^{n} v_i^2}$$

Notice that $||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$, so that the norm inherits the following properties from the scalar product, for every $\alpha \in \mathbb{R}$ and $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$:

1. **Positive homogeneity**: $||\alpha \mathbf{v}|| = |\alpha| ||\mathbf{v}||$.
2. **Triangular inequality**: $||\mathbf{v} + \mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||$.
3. **Positive Definite**: $||\mathbf{v}|| \geq 0$ and $||\mathbf{v}|| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

The following is a fundamental inequality relating norms and scalar products:

Theorem 1 (Cauchy-Schwarz inequality)**.** *Let* $\mathbf{v}$ *and* $\mathbf{w}$ *be two vectors in* $\mathbb{R}^n$*. Then*

$$|\langle \mathbf{v}, \mathbf{w} \rangle| \leq ||\mathbf{v}|| \cdot ||\mathbf{w}||$$

with equality if and only if $\mathbf{v} = \alpha \mathbf{w}$ *for some* $\alpha \in \mathbb{R}$.

The Euclidean norm of a vector measures the distance of the vector coordinates from the origin. On the other hand, the scalar product between two vectors gives information about the angle between them. In particular, we can define the cosine between two vectors in the following way: let $\mathbf{v}$ and $\mathbf{w}$ be two vectors in $\mathbb{R}^n$. Then

$$\cos(\angle \mathbf{vw}) = \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{||\mathbf{v}||||\mathbf{w}||}$$

In $\mathbb{R}^2$ and $\mathbb{R}^3$, the definition of cosine matches the geometric intuition, as the cosine of the angle between the two vectors will be exactly that of the definition. Notice that by the Cauchy-Schwarz inequality the absolute value of $\cos(\angle \mathbf{vw})$ will always be bounded above by one.

Example: Let $\mathbf{v} = \begin{bmatrix} 0 & 3 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 1 & 1 \end{bmatrix}$. Consider the vectors in $\mathbb{R}^2$ as the arrows from the origin to the corresponding coordinates. Then $||\mathbf{v}||$ and $||\mathbf{w}||$ measure the length of these arrows, or equivalently, the distance of the points from the origin. In particular:

$$||\mathbf{v}|| = \sqrt{0 \cdot 0 + 3 \cdot 3} = 3, \quad ||\mathbf{w}|| = \sqrt{1 \cdot 1 + 1 \cdot 1} = \sqrt{2}$$

and

$$\langle \mathbf{v}, \mathbf{w} \rangle = 0 \cdot 1 + 3 \cdot 1 = 3$$
$$\cos(\angle \mathbf{vw}) = \frac{3}{3 \cdot \sqrt{2}} = \frac{1}{\sqrt{2}}$$

Notice that $\cos(\angle \mathbf{vw}) = \cos(\frac{\pi}{4})$, which is exactly the angle spanned by the two vectors, as in Figure 3.

Matrices and their linear operations

A matrix is a list of real numbers whose position is identified by two indices: the row index and the column index. It is usually represented as

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0.5 & 0 \end{bmatrix}$$

A matrix does not need to have the same number of rows and columns (in this case, the matrix has 2 rows and 3 columns). If it has, it's called a *square* matrix. A matrix A is generally an element of $\mathbb{R}^{n \times m}$, where n is the number of rows, while m is the number of columns. Each row can be considered as a separate $1 \times m$ vector, while each column is a $n \times 1$ vector. In particular, if $\mathbf{a}_1, ..., \mathbf{a}_m$ are vectors of the same dimension n, we can write

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & ... & \mathbf{a}_m \end{bmatrix}$$

to represent the matrix having $\mathbf{a}_i$ as i-th column vector. Analogously, if $\mathbf{b}_1, ..., \mathbf{b}_n$ are m-dimensional row vectors, we can write

$$B = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ ... \\ \mathbf{b}_n \end{bmatrix}$$

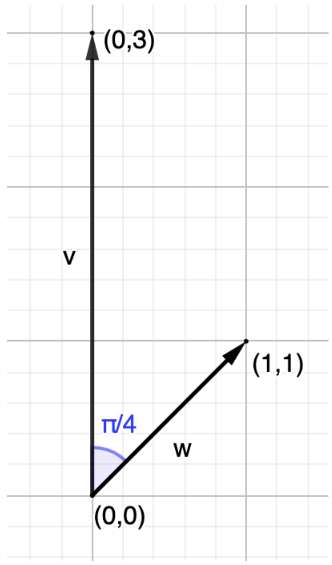

Figure 3: Angle between vectors $\mathbf{v} = \begin{bmatrix} 0 & 3 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 1 & 1 \end{bmatrix}$

as the matrix B having $\mathbf{b}_i$ as i-th row.

Some terminology for square matrices of dimension n: the elements $\{A_{i,i}\}_{i=1}^n$ are the elements in the (principal) diagonal of A. A is a *diagonal matrix* if the only nonzero elements are on the diagonal. It is *upper-triangular* if its only nonzero elements are on and above the diagonal, and it is *lower-triangular* if the only nonzero elements are on and below the diagonal.

$$\begin{bmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{diagonal matrix.}$$

$$\begin{bmatrix} 3 & 2 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 9 \end{bmatrix} \qquad \text{upper-triangular matrix.}$$

$$\begin{bmatrix} 3 & 0 & 0 \\ -1 & -1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \qquad \text{lower-triangular matrix.}$$

The first basic operation on matrix that we introduce is *transposition*: Let A be a $n \times m$ matrix A, then the matrix A^T is called the *transpose of* A if

$$A^T_{i,j} = A_{j,i} \qquad \text{for every } i = 1, ..., n \quad \text{and} \quad j = 1, ..., m.$$

In particular, A^T will be an $m \times n$ matrix. For example:

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0.5 & 0 \end{bmatrix}^T = \begin{bmatrix} 1 & 2 \\ 0 & 0.5 \\ -1 & 0 \end{bmatrix}$$

Notice that by transposing row vectors are transformed into column vectors and viceversa. Moreover, transposition is an *involution*, that is $(A^T)^T = A$. Finally, if A is a square matrix and $A^T = A$ we say that A is *symmetric*.

The second operation we introduce is *addition* or *sum* between matrices. This is only possible when two matrices have the same dimension, in which case the sum of two $n \times m$ matrices A and B is a matrix $A + B$ whose (i, j)-the component is the sum of the respective components of A and B:

$$(A + B)_{i,j} = A_{i,j} + B_{i,j}$$

for all $i = 1, ..., n$ and $j = 1, ..., m$. For example,

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0.5 & 0 \end{bmatrix} + \begin{bmatrix} 0 & -0.1 & 1 \\ 3 & 1.5 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -0.1 & 0 \\ 5 & 2 & 0 \end{bmatrix}$$

Matrix addition has similar properties to vector addition. These are, for $A, B, C \in \mathbb{R}^{n \times m}$:

1. **Associativity**: $(A + B) + C = A + (B + C)$.
2. **Commutativity**: $A + B = B + A$.
3. **Existence of Identity element**: if 0 is a $n \times m$ -dimensional matrix of zeroes, $0 + A = A$.
4. **Existence of inverse** $A + (-A) = 0$, where $(-A)$ is the matrix such that $(-A)_{i,j} = -A_{i,j}$.
5. **Transpose of the sum**: $(A + B)^T = A^T + B^T$

The third operation we introduce is *scalar multiplication of a matrix*, which for a given scalar α and $n \times m$ matrix A gives a $n \times m$ matrix αA such that

$$(\alpha A)_{i,j} = \alpha A_{i,j}$$

For example:

$$-2 \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0.5 & 0 \end{bmatrix} = \begin{bmatrix} -2 & 0 & 2 \\ -4 & -1 & 0 \end{bmatrix}$$

The properties are the same as scalar multiplication by vectors, that is, for every $A, B \in \mathbb{R}^{n \times m}$ and $\alpha, \beta \in \mathbb{R}$:

1. **Compatibility with field multiplication**: $(\alpha\beta)A = \alpha(\beta A)$.
2. **Existence of identity element**: $1A = A$ for every A.

3. **Distributivity with respect to vector addition**: $\alpha(A+B) = \alpha A+\alpha B$.

4. **Distributivity with respect to scalar addition**: $(\alpha+\beta)A = \alpha A+\beta A$.

Notice that these properties imply that the space of all $n \times m$ matrices with these two operations is a vector space, as in the previous definition.

Matrix-matrix and matrix-vector multiplication

A fundamental operation we need to consider is *matrix multiplication*. In general, multiplication between matrices A and B is well defined only if the number of rows of B is equal to the number of columns of A. In particular, if A is a $n \times m$ matrix and B is a $m \times l$ matrix, then AB is a $n \times l$ matrix which, component-wise, is defined as follows:

$$(AB)_{i,j} = \sum_{k=1}^{m} A_{i,k} B_{k,j}$$

This definition reminds us of the scalar product. In fact, we can say that the (i, j)-th element of the matrix AB is obtained as a scalar product of the i-th row vector of A with the j-th column vector of B, which is well-defined because they have the same size. For example:

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0.5 & 0 \end{bmatrix} \begin{bmatrix} 0.5 & 1 \\ 0 & 2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} -0.5 & 2 \\ 1 & 3 \end{bmatrix}$$

In this case, we multiply a 2×3 matrix by a 3×2 matrix, thus obtaining a 2×2 matrix. The element $(AB)_{1,1}$, for instance, is obtained by calculating the scalar product between the first row of A with the first column of B, that is

$$\langle \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}, \begin{bmatrix} 0.5 \\ 0 \\ 1 \end{bmatrix} \rangle = 1 \cdot 0.5 + 0 \cdot 0 + (-1) \cdot 1 = -0.5$$

All the other components are obtained in the same way.

Vectors can be understood as a particular case of matrices in which either the row or the column have dimension 1. In particular, if $\mathbf{b}$ is a $m \times 1$ vector, $A\mathbf{b}$ is a $n \times 1$ vector. In this particular instance, we call the operation *matrix-vector multiplication*.

Notice that matrix multiplication is in general **NOT commutative**, that is $AB \neq BA$, even if A and B are both square matrices of the same dimension. For example, consider

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

while

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

Generally if multiplication is well-defined for A, B and C, the following properties hold:

1. **Associativity**: $(AB)C = A(BC)$.
2. **Right distributivity with respect to addition**: $(A+B)C = AC+AB$.
3. **Left distributivity with respect to addition**: $A(B+C) = AB+AC$.
4. **Associativity with respect to scalar multiplication**: $\alpha(AB) = (\alpha A)B = A(B\alpha)$.
5. **Transpose of the product**: $(AB)^T = B^T A^T$

In particular, square matrices are closed under matrix multiplications. That means that if A, B are in $\mathbb{R}^{n\times n}$, so is AB. In this case we have the additional property:

6. **Existence of identity element**: $AI = IA = A$ for I a diagonal $n \times n$ matrix of 1.

The multiplicative inverse of a matrix is generally not well-defined, unless some properties are satisfied. We will talk more about inverses in future lectures after defining determinants.

Computational Complexity and sparse matrices

We have given formal definition on a few operations involving matrices and vectors. For the purpose of this class we turn to the question of what happens when these operations are performed by a machine. In this case, real numbers are stored using *floating point format*, so that every number can be represented with a precision of about 10^{-10} (tenth decimal). Every time there is an elementary operation (addition, subtraction, multiplication or division), the machine rounds the result to the closest floating point number, introducing a rounding error. While in general such error is not relevant, it is important to understand how many times the value is rounded both to minimize the error and to minimize the time required for complex operations.

In what follows, we will call addition, subtraction, multiplication and division between two scalars a *floating point operation* or *flop*. The number of flops needed to carry an operation is called the *complexity* of such operation. For example, in order to compute the sum between two vectors in $\mathbb{R}^n$, we need n flops (one for each coordinate). Scalar multiplication also has the same complexity, since each component of the vector has to be multiplied by the same scalar.

The scalar product between two vectors has complexity $2n-1$, since we first have to multiply each pair of coordinate and then they need to be added to a running sum. Matrix multiplication can be seen as multiple scalar products, one for each component of the resulting matrix. Therefore, if we are multiplying a $m \times n$ matrix by a $n \times l$ matrix, we will need $ml(2n-1)$ flops.

Sometimes the same operation can be computed in different ways, thus changing the number of required flops. For instance, consider A, B matrices in $\mathbb{R}^{n\times n}$ and $\mathbf{x}, \mathbf{y}$ vectors in $\mathbb{R}^n$. The operation $(A+B)(\mathbf{x}+\mathbf{y})$ (computed in the order described by the parenthesis) requires n^2 flops for $(A+B)$, n flops for $(\mathbf{x}+\mathbf{y})$ and then n scalar products between each row of $(A+B)$ and the column vector $(\mathbf{x}+\mathbf{y})$. Since each scalar product has complexity $2n-1$, the total number of flops is

$$tot = n^2 + n + n(2n-1) = 3n^2$$

On the other hand, if we compute $A\mathbf{x} + A\mathbf{y} + B\mathbf{x} + B\mathbf{y}$, we will compute 4 matrix-vector multiplications, each having complexity $n(2n-1)$ and then we will sum all the components. The total number of flops is therefore

$$tot = 4n(2n-1) + 3n = 8n^2 - n$$

In general, we look at the leading term and choose the one with the smallest complexity. Therefore computing $(A+B)(\mathbf{x}+\mathbf{y})$ is faster than computing $A\mathbf{x} + A\mathbf{y} + B\mathbf{x} + B\mathbf{y}$.

Another important tool in terms of computational complexity is given by **sparse vectors and matrices**. A matrix (or a vector) is called *sparse* if it contains a large number of zeros. On such objects it makes sense to define a function $\mathbf{nnz}(x)$ returning the number of nonzero components. Whenever one of the elements of a basic operation (addition, multiplication, subtraction, division) is zero, there is no floating point approximation, so that the result is not counted in the flops. This introduces some advantages when dealing with sparse matrices as for instance when considering the sum between two sparse vector $\mathbf{x}$ and $\mathbf{y}$, the complexity will be $\min\{\mathbf{nnz}(\mathbf{x}), \mathbf{nnz}(\mathbf{y})\}$. At the same time, the scalar product between $\mathbf{x}$ and $\mathbf{y}$ will have complexity equal to $2\min\{\mathbf{nnz}(\mathbf{x}), \mathbf{nnz}(\mathbf{y})\}$.