Lecture 1: Generalities on vectors and matrices

Francesco Preta
07/07/2020

What is a vector?

Vectors are 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 $\mathbb{R}^n$.

Vectors

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

Possible interpretations of vectors

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.

Sum of vectors

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

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

or, in coordinate notation:

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

Example:

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

Properties of vector addition:

For every $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$:

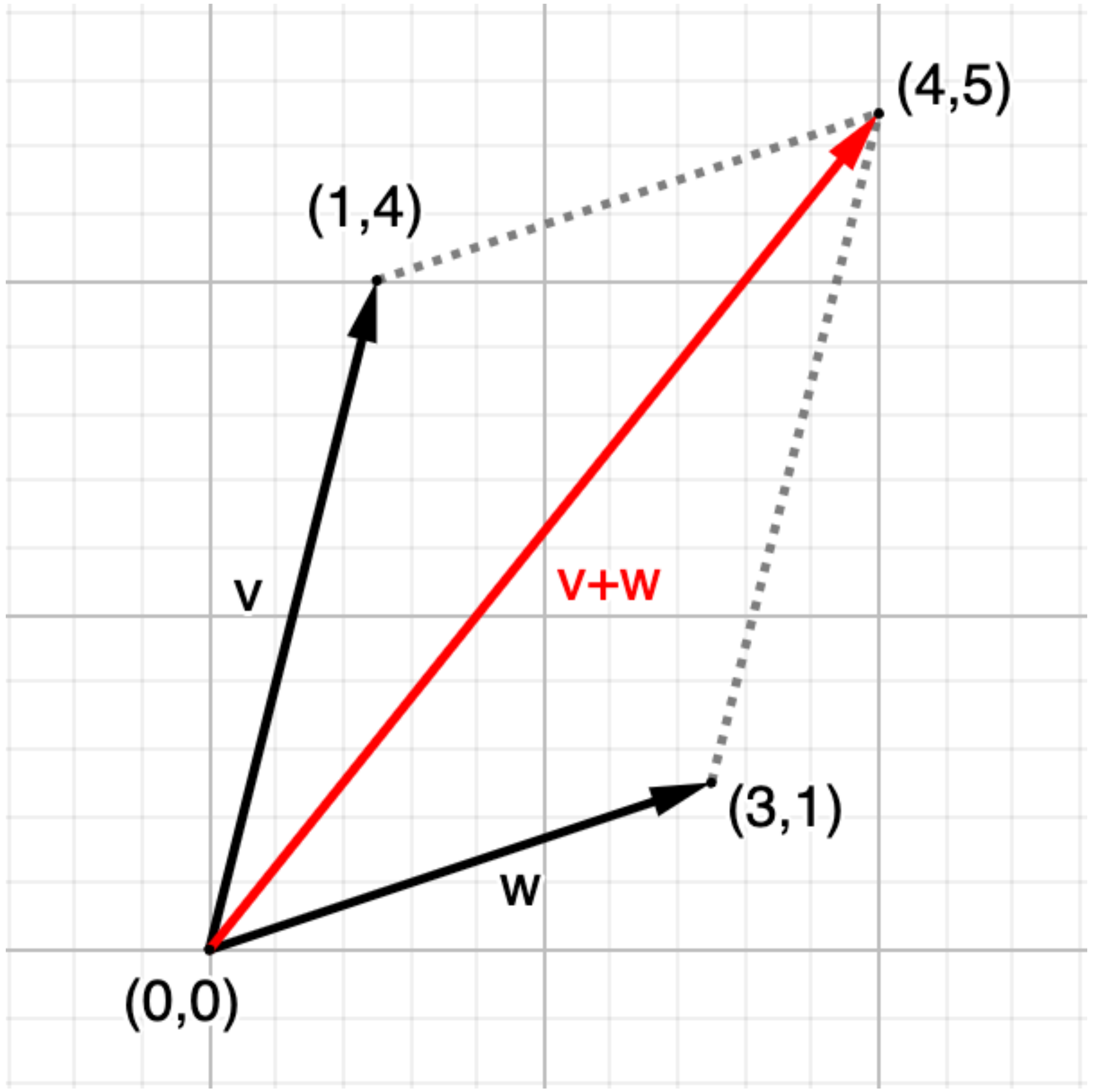
- **Associativity**: $(\mathbf{v} + \mathbf{w}) + \mathbf{u} = \mathbf{v} + (\mathbf{w} + \mathbf{u})$.
- **Commutativity**: $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$.
- **Existence of Identity element**: if $\mathbf{0}$ is a n-dimensional vector of zeroes,$\mathbf{0} + \mathbf{v} = \mathbf{v}$.
- **Existence of inverse**: $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$ where $-\mathbf{v}$ is the vector such that $(-\mathbf{v})_i = -\mathbf{v}_i$.

The parallelogram rule

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.

Geometric interpretation:

$$\begin{bmatrix} 1 \\ 4 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$



Scalar multiplication of a vector

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

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

Example:

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

Properties of scalar multiplication:

for every $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$:

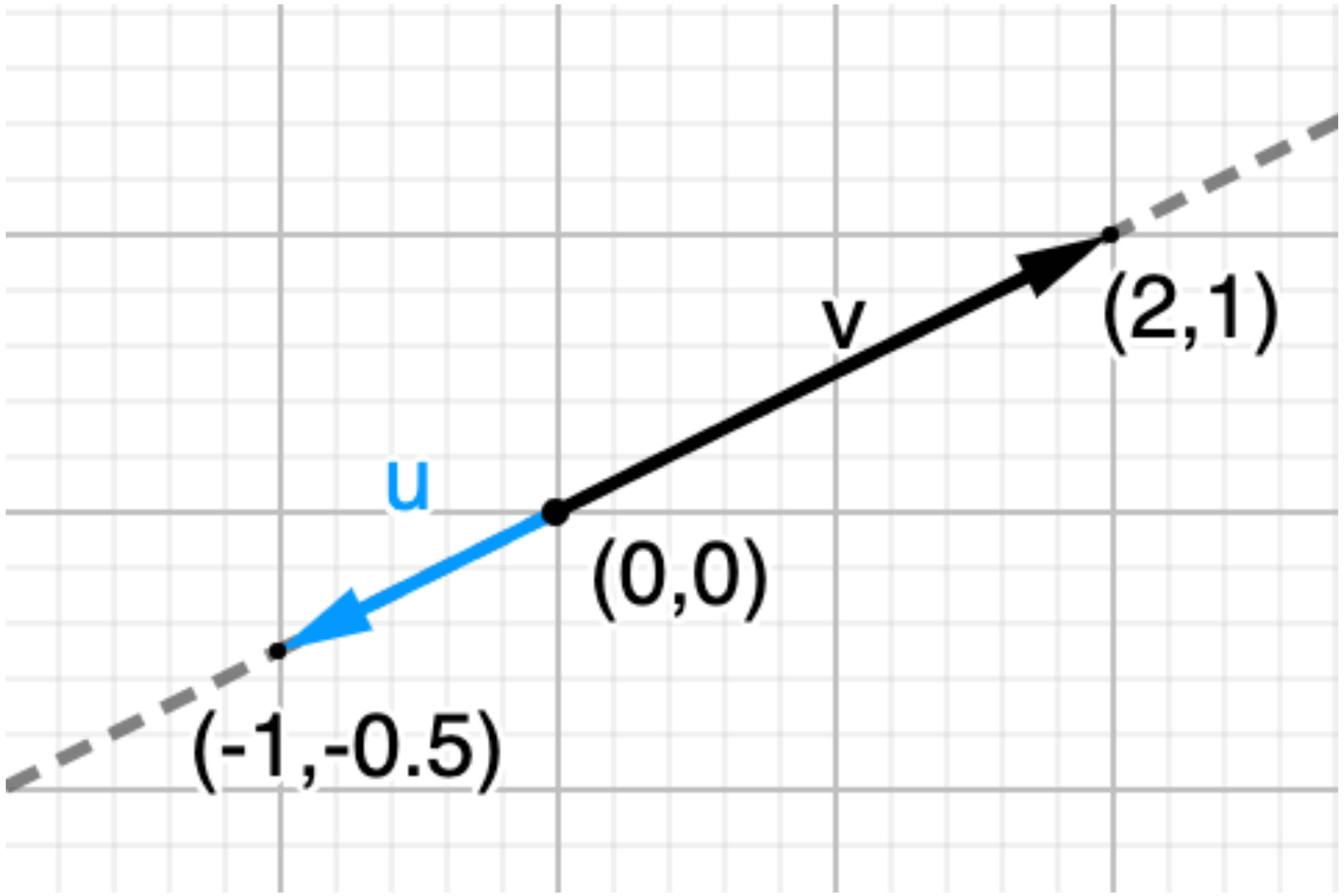
- **Compatibility with field multiplication**: $(\alpha\beta)\mathbf{v} = \alpha(\beta\mathbf{v})$.
- **Existence of identity element**: $1\mathbf{v} = \mathbf{v}$.
- **Distributivity with respect to vector addition**: $\alpha(\mathbf{v} + \mathbf{w}) = \alpha\mathbf{v} + \alpha\mathbf{w}$.
- **Distributivity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

Geometric interpretation:

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.



Definition: linear combination

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

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

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

Example:

$\mathbf{v}_1 = [1 \quad 2 \quad 0], \mathbf{v}_2 = [-1 \quad 1 \quad 0]$

Definition: span

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

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

The canonical basis

In $\mathbb{R}^n$ we can define a set $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ 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 $\mathbb{R}^n = \text{span}(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ since for every $\mathbf{x} \in \mathbb{R}^n$

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

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

Abstract vector spaces

Definition: vector space

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:

- **Associativity of +**: $(\mathbf{v} + \mathbf{w}) + \mathbf{u} = \mathbf{v} + (\mathbf{w} + \mathbf{u})$.
- **Commutativity of +**: $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$.
- **Existence of Identity element for +**: there exists an element $\mathbf{0}$ such that $\mathbf{0} + \mathbf{v} = \mathbf{v}$.
- **Existence of inverse for +**: for every $\mathbf{v} \in V$ there exists an element $\mathbf{v}'$ (called the inverse of $\mathbf{v}$) such that $\mathbf{v} + (\mathbf{v}') = \mathbf{0}$.

(continues on the next page)

Definition: vector space (continued)

- **Compatibility of $\cdot$ with field multiplication**: $(\alpha\beta) \cdot \mathbf{v} = \alpha(\beta \cdot \mathbf{v})$.
- **Existence of identity element for $\cdot$**: $1 \cdot \mathbf{v} = \mathbf{v}$.
- **Distributivity of $\cdot$ with respect to +**: $\alpha \cdot (\mathbf{v} + \mathbf{w}) = \alpha \cdot \mathbf{v} + \alpha \cdot \mathbf{w}$.
- **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 $\alpha, \beta \in \mathbb{R}$,
$\alpha\mathbf{v} + \beta\mathbf{w} \in V$.

Meaning of the definition:

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:

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

Example 2:

span$(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ for $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset \mathbb{R}^n$

Scalar product, norm and cosine

Definition: scalar product

For any pair of vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$ their **scalar product** is defined as

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

Example:

$$\left\langle \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} \right\rangle$$

Properties:

For every $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$, the scalar product has the following properties:

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

Definition: Euclidean norm

The **Euclidean norm** of a vector $\mathbf{v}$ in $\mathbb{R}^n$ is the quantity

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

Properties:

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 $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$:

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

Cauchy-Schwarz inequality

Theorem

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 $\alpha \in \mathbb{R}$.

Geometric meaning and cosine between vectors

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

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

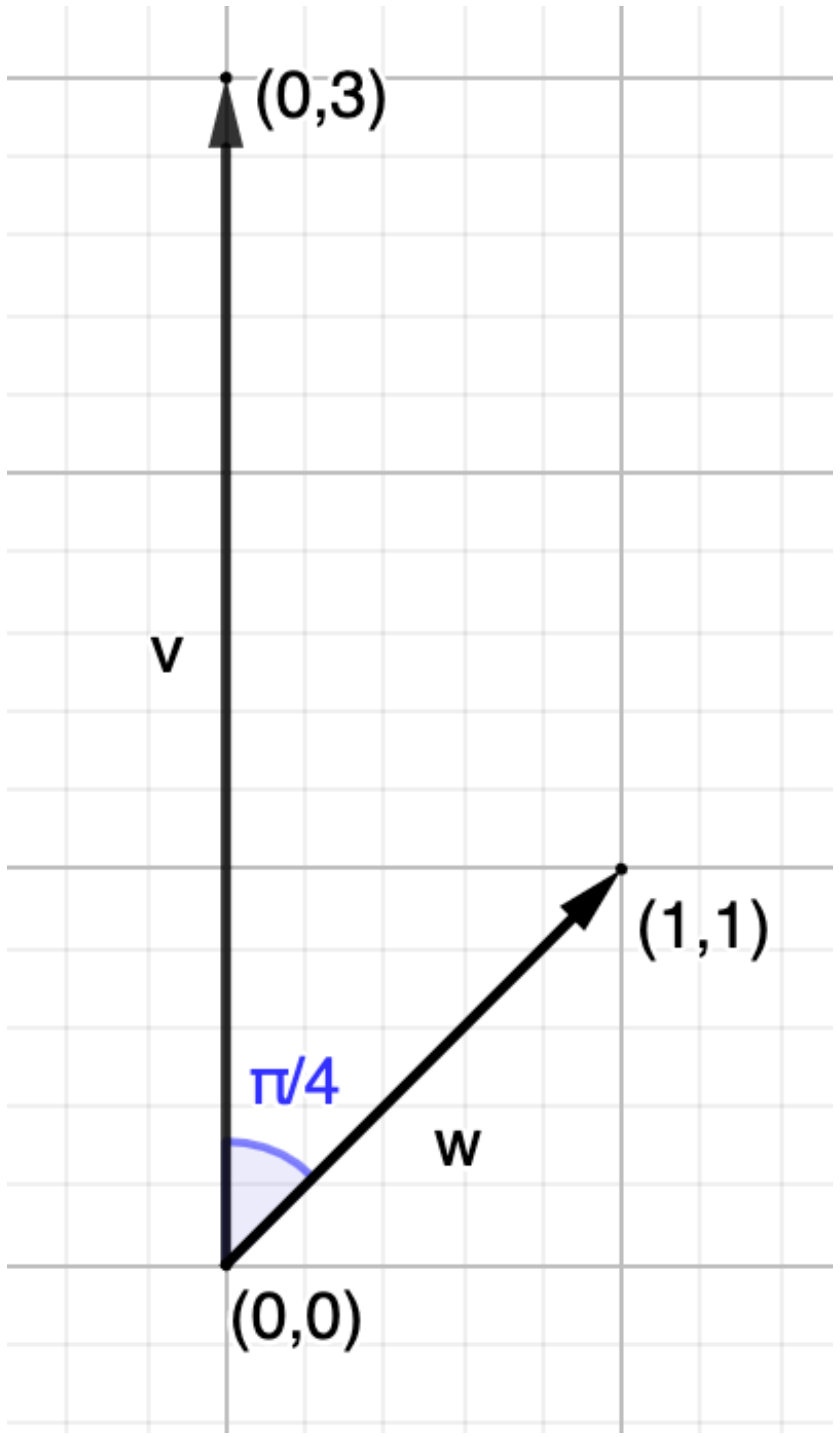
Geometric intuition

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 $\cos(\angle \mathbf{v}\mathbf{w})$ will always be bounded above by one.

Example:

$\mathbf{v} = [0 \quad 3], \mathbf{w} = [1 \quad 1]$

Example: (continued)



Matrices and their linear operations

What is a matrix?

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 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.

Defining matrices by column or row vectors

If $A \in \mathbb{R}^{n \times m}$ Each row can be considered as a separate $1 \times m$ vector, while each column is a $n \times 1$ vector. In particular, if $\mathbf{a}_1, \ldots, \mathbf{a}_m$ are vectors of the same dimension n, we can write

$$A = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \ldots \quad \mathbf{a}_m]$$

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

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

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

Square matrices

If a matrix has the same number of rows and columns, it's called a **square** matrix.

For a square matrix, 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.

Examples:

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

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

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

Definition: transpose of a matrix

Let A be a $n \times m$ matrix A, then the matrix A^T is called the **transpose of** A if

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

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

Example:

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

Symmetric matrices

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**.

Matrix addition

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)-th component is the sum of the respective components of A and B:

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

for all $i = 1,\ldots,n$ and $j = 1,\ldots,m$.

Example:

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

Properties:

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

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

Scalar multiplication of a matrix

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

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

Example:

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

Properties:

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

- **Compatibility with field multiplication**: $(\alpha\beta)A = \alpha(\beta A)$.
- **Existence of identity element**: $1 \cdot A = A$ for every A.
- **Distributivity with respect to vector addition**: $\alpha(A + B) = \alpha A + \alpha B$.
- **Distributivity with respect to scalar addition**: $(\alpha + \beta)A = \alpha A + \beta A$.

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

Matrix 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:

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

Matrix multiplication and scalar product

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.

Example:

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

Matrix-vector multiplication

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**.

Non-commutativity of matrix multiplication

Notice that matrix multiplication is in general **NOT commutative**, that is $AB \neq BA$, even if both products are well defined. For example,

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

while

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

Properties

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

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

Properties for square matrices:

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:

- **Existence of identity element**: $AI = IA = A$ for I an $n \times n$ diagonal matrix having only 1 on the diagonal.

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

Floating point format

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}.

Rounding errors

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.

Flops

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.

Examples: vector sum and scalar multiplication

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.

Examples: scalar product and matrix multiplication

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.

Example: optimizing the flop count

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 $\mathbb{R}^n$. Compare the operations:

- $(A + B)(\mathbf{x} + \mathbf{y})$
- $A\mathbf{x} + A\mathbf{y} + B\mathbf{x} + B\mathbf{y}$

Sparse vectors and matrices

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.

Flop count for sparse objects

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.

Examples: sparse vector sum and scalar product

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 $2\min\{\mathbf{nnz}(\mathbf{x}), \mathbf{nnz}(\mathbf{y})\}$.