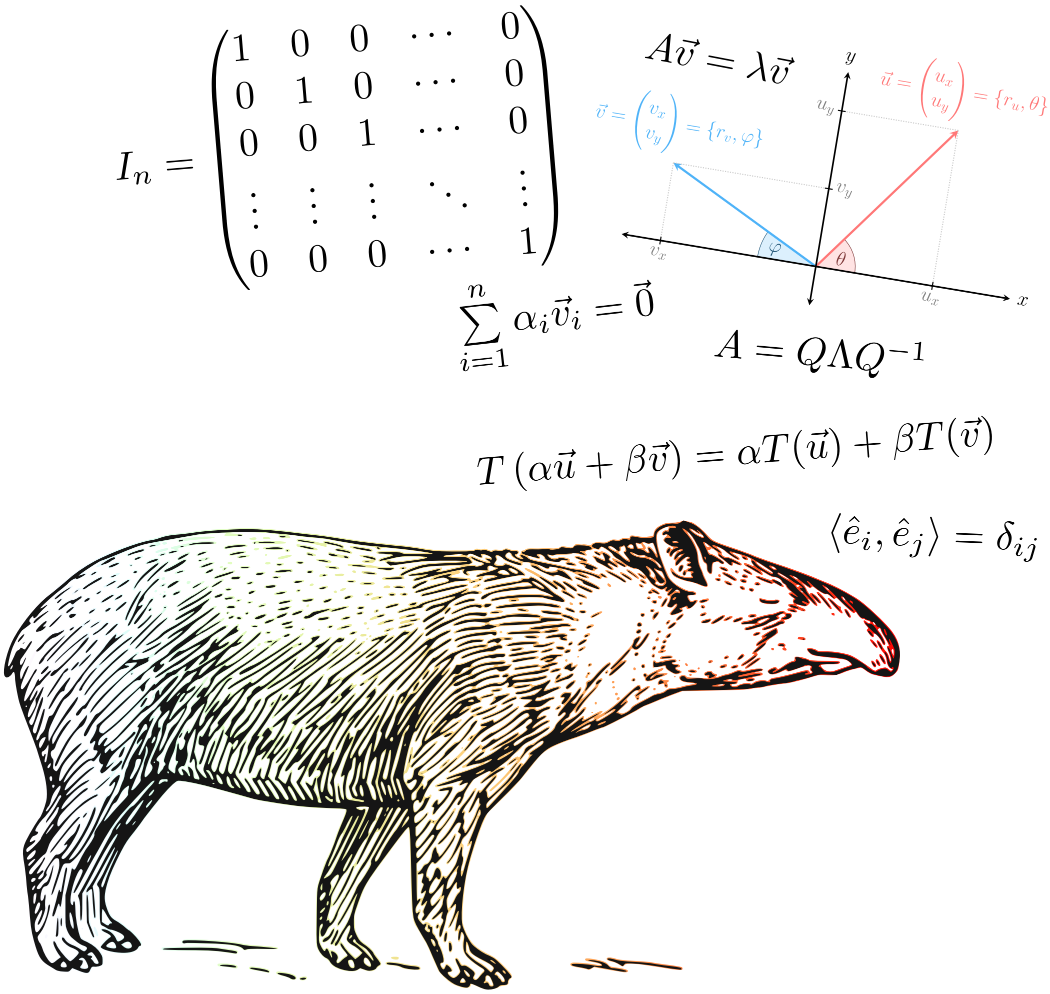


Basic Linear Algebra for Non-Mathematicians

Lecture Notes

Peleg Bar Sapir

March 30, 2020

Preface 7

1. Introduction 9

1.1. Mathematical Logic 9

1.1.1. Propositions 9

1.2. Sets 10

1.2.1. Basic Properties 10

1.2.2. Subsets and Supersets 12

1.2.3. Intersections and Unions 12

1.2.4. Difference of Sets 14

1.2.5. Complement 14

1.2.6. Power Set 15

1.2.7. Important Number Sets 15

1.2.8. Cartesian Products 16

1.3. Functions 17

1.3.1. Relations 17

1.3.2. Basic Definition of a Function 18

1.3.3. Domain and Image of a Function 19

1.3.4. Injective, Surjective and Bijective Functions 19

1.3.5. Multivariable Functions 22

1.4. Graphs 23

2. Vectors 27

2.1. Introduction 27

2.2. Geometric Approach 27

2.2.1. Basics 27

2.2.2. Scaling a Vector, Adding Vectors 28

2.3. Algebraic Approach 31

2.3.1. Basics 31

2.3.2. Scaling a Vector, Adding Vectors 34

2.4. Norm (Length) and Argument (Angle) of a Vector 36

2.5. Spaces and Subspaces 38

2.5.1. Linear Dependency and Linear Combinations 38

2.5.2. Vector Spaces and Bases 40

2.6. Vector-Vector Products . . . 43
2.6.1. The Dot Product . . . 43
2.6.2. The Cross Product . . . 45

3. Linear Transformations **49**
3.1. What is a Linear Transformation? . . . 49
3.2. Transforming Vectors . . . 50
3.3. Transforming Spaces . . . 51
3.4. Characteristics of Linear Transformations . . . 52
3.5. Basic Types of Linear Trasformations . . . 53
3.5.1. Rotations Around the Origin . . . 53
3.5.2. Scaling . . . 53

4. Matrices **57**
4.1. Matrices from Linear Transformations . . . 57
4.2. Basics of Matrices . . . 58
4.3. Diagonal and Triangular Matrices . . . 60
4.4. Transpose of a Matrix . . . 60
4.5. Adding Two Matrices . . . 61
4.6. Multiplying a Matrix by a Scalar . . . 62
4.7. Multiplying a Matrix and a Vector . . . 62
4.8. Important Matrices . . . 64
4.8.1. The Identity Matrix . . . 64
4.8.2. Rotation Matrices . . . 64
4.8.3. Scaling Matrices . . . 65
4.8.4. Reflection Matrices . . . 66
4.8.5. Skew Matrices . . . 66
4.9. Multiplying Two Matrices . . . 67
4.10. Matrix-Matrix Multiplication as Transformation Composition . . . 69
4.11. Trace of a Matrix . . . 70
4.12. The Determinant . . . 70
4.12.1. Negative Determinant . . . 71
4.12.2. Zero Determinant . . . 72
4.12.3. Calculating the Determinant . . . 73
4.12.4. Determinant of Matrix-Matrix Product . . . 75
4.13. Inverse Transformations . . . 75
4.13.1. Inverse Matrices . . . 75
4.13.2. Kernel and Nullspace of a Transformation . . . 76

5. Systems of Linear Equations **79**

5.1. Basics . . . 79

5.2. Solution Set . . . 80

5.2.1. Geometric Interpretation of the Solution Set . . . 80

5.2.2. General Number of Solutions . . . 81

5.3. Solving a System of Linear Equations . . . 82

5.3.1. Eliminating Variables . . . 82

5.3.2. Row Reduction (Gaussian Elimination) . . . 82

5.3.3. Underdetermined Row Echelon Matrix . . . 83

6. Eigenvectors and Eigenvalues **85**

6.1. Definition . . . 85

6.2. Characteristic Polynomial . . . 86

6.3. Eigenspaces, Multiplicity . . . 88

6.4. Matrix Diagonalization . . . 89

6.5. Summary of Eigenvectors and Eigenvalues in $\mathbb{R}^2$ . . . 90

7. Real-Life Uses of Linear Algebra **91**

7.1. Stable Populations in Predator-Pray Models / Leslie Models? . . . 91

7.2. Least Squares Approximation . . . 93

7.2.1. Linear Functions . . . 93

7.2.2. General Polynomials . . . 96

7.2.3. Other Functions . . . 97

7.3. Principal Component Analysis (PCA) . . . 98

8. General Vector Spaces **99**

8.1. $\mathbb{R}^n$ as a Vector Space . . . 99

8.2. Examples of Vector Spaces . . . 101

8.2.1. Polynomials of Degree $\leq \mathbf{n}$ . . . 101

8.2.2. Real Valued Functions . . . 103

This booklet contains the lecture notes for part I of a course given in the summer semester of 2020, at the Georg-August-Universität in Göttingen, Germany. The course in question is called *Mathematics and Computer Science (B.MES.108)*, and was given as part of the bachelor program *Molecular Ecosystem Sciences* of the *Faculty of Forest Sciences and Forest Ecology*, by the *Department Ecoinformatics, Biometrics and Forest Growth.*

The course is intended on introducing the students, in their first year of studies, to basic mathematical tools that they would use in the course of their scientific career. **It is not intended to provide in-depth mathematical content**, but rather give a general overview of topics such as Linear algebra (given here), analysis of real functions and computer science. As such, the lectures do not dive deep into any particular topic, and are not to be used as an exhaustive reference on the matter (especially not from a mathematician's point of view, as they provide a more applicable approach rather than a pure theoretical one).

In this booklet, the reader will find notes on the different topics covered in the first part of the lectures. These notes are combined with practical examples, signified by a pale blue bounding box, such as follows:

Example 1

This is a practical example.

In addition, two color-coded ”boxes” can also be found: the first is a ”warning”/”attention” box which stresses important notes and possible pitfalls to be aware of:

Note 1

This is an attention box.

The second is a ”challange” box, which challenges the reader to do some calculation by themselves, or to prove a lemma/theorem. These challenges are not stricly mandatory for understanding and/or successfuly studying the topic; it is however highly adviced to at least try them out.

Challange 1

This is a challange box.

In addition, some other boxes are used troughout the booklet more sparingly, such as a ”definition” box, and a box that signifies extremely important concepts.

If you find any mistakes in the text (including typos and misspelling), please contact me per email, and let me know.

Lastly, unless otherwise mentioned, all graphics in this booklet were created by me, using the Ti*k*Z package, the Ipe extensible drawing editor and Inkscape. The typesetting was done in LaTeX.

Introduction 1

1.1 Mathematical Logic

1.1.1 Propositions

Propositions are statements that can be either **true** or false.

Example 2

- The moon is smaller than the Earth (**true**)
- $1 + 3 = 4$ (**true**)
- Protons have no electric charge (false)
- $12 > 13$ (false)

Propositions can be grouped together with *operators* such as **and**, **or**. The **and** operator returns a **true** statement only if *both* the statements it groups are themselves true, otherwise it returns false.

Example 3

Combining statements using the **and** operator (statememnts highlighted in blue are true, statements highlighted in red are false):

$$\boxed{1+2=3} \text{ and } \boxed{3-5=-2} \Rightarrow \text{true}$$
$$\boxed{1+2=3} \text{ and } \boxed{2\times 4=7} \Rightarrow \text{false}$$
$$\boxed{\frac{10}{2}=1} \text{ and } \boxed{2^4=16} \Rightarrow \text{false}$$
$$\boxed{7<5} \text{ and } \boxed{10+2=13} \Rightarrow \text{false}$$

The **or** operator returns **true** if **at least** one of the statements it groups is true.

Example 4

Combining statements using the **or** operator (statememnts highlighted in blue are true, statements highlighted in red are false):

$$\boxed{1+2=3} \text{ or } \boxed{3>7} \Rightarrow \text{true}$$
$$\boxed{0+3=-1} \text{ or } \boxed{1=1} \Rightarrow \text{true}$$
$$\boxed{2\times 2=4} \text{ or } \boxed{2+0=2} \Rightarrow \text{true}$$
$$\boxed{3\times 7=10} \text{ or } \boxed{\frac{1}{2}<\frac{1}{10}} \Rightarrow \text{false}$$

We can summarize this in a *truth table*, for combinations of statements A and B, which could each be either true or false:

Truth Table for the Operators AND, OR

A	B	AND	OR
true	true	true	true
true	false	false	true
false	true	false	true
false	false	false	false

The operators **and**, **or** and other operators have mathematical abbreviations (which are called *notation*):

Common Mathematical Operators

Symbol	In words	Comments
$\neg a$	**not** a	"Flips" true to false, and false to true.
$a \wedge b$	a **and** b	
$a \vee b$	a **or** b	
$a \Rightarrow b$	a **implies** b	"if a is true, then b is also true"
$a \Leftrightarrow b$	a **is equivalent to** b	"if and only if a is true, then b is true".
$\forall x$	**For all** x (...)	
$\exists x$	**There exists** x **such that** (...)	
$a := b$	a **is defined to be** b	

1.2 Sets

1.2.1 Basic Properties

A set is a collection of *elements*. Elements of a set can be any concept - be it physical (a chair, a car, a tapir) or abstract (a number, an idea). In fact, elements of a set can themselves be sets. In this course we limit ourselves to only sets of numbers and other mathematical objects. Sets can have a *finite* number of elements (for example, the set of all people) - or an *infinite* number of elements (for example, the set of all numbers bigger than 3).

We use curly brackets for notating sets.

Example 5

$$\{1, 2, 3, 4\}, \left\{-3, 7, \pi, 0.1, 1337, \frac{1}{17}\right\}$$

The order of elements in a set does not matter.

Example 6

The follows sets are all equal:

$$\{1, 2, 3, 4\} = \{1, 3, 2, 4\} = \{4, 1, 2, 3\}$$

There are no repetitions in a set.

Example 7

$$\{1, 2, 3, 3, 3, 4\} = \{1, 2, 3, 4\}.$$

Sets are usually designated by upper-case letters, while their elements are usually designated with lower-case letters.

Example 8

$$A = \{0, 3, -1\}, B = \left\{-7, \pi, 0, \frac{1}{2}\right\}$$

The symbol $\in$ means that an element is in a set, and $\notin$ means that an element is not in a set.

Example 9

For the sets A, B defined above:

$$3 \in A,\ 4 \notin A,\ \pi \in B,\ 1 \notin B$$

We can also define sets using a proposition.

Example 10

$$M = \left\{x \mid x > 1 \text{ and } x < 3\right\}$$

The vertical line separator $|$ means "such that", and so the set M is defined as "x such that x is bigger than 1 and smaller than 3". An alternative notation would be

$$M = \left\{x \mid 1 < x < 3\right\}$$

Note 2

For the above set, $1 \notin M$ and also $3 \notin M$. This is because of the use of $<$ (*smaller than*), and not $\leq$ (*smaller than or equal to*).

The number of elements in a set (also called its *cardinality*) is notated with two vertical bars.

Example 11

$$M = \{1, 4, 6, 13, 17\} \Rightarrow |M| = 5$$

A special set is the *empty set*, notated as $\emptyset$. It has the unique property that $\left|\emptyset\right| = 0$.

1.2.2 Subsets and Supersets

If A contains all elements of B (and perhaps other elements as well), then A is a *superset* of B, and B is a *subset* of A.

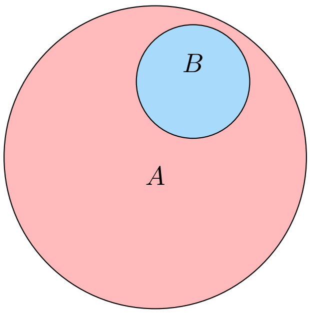
Example 12

The sets $\{0, -3\}, \{2, -5, 7\}, \{1, -3, 0\}$ are all subsets of $\{1, 2, -5, -3, 0, 7\}$. (these are not **all** the subsets of $\{1, 2, -5, -3, 0, 7\}$, only three examples)

The superset/subset relation between A and B is written as

$$B \subseteq A$$

Using a *Venn diagram*, we could visualize this as follows:



If for some two sets M, N both $M \subseteq N$ and $N \subseteq M$, then both sets are identical. Writing this formally:

$$M \subseteq N \wedge N \subseteq M \Leftrightarrow N = M$$

1.2.3 Intersections and Unions

We define the *intersection* of two sets A, B as the set of all elements that are *both* in A **and** in B.

Example 13

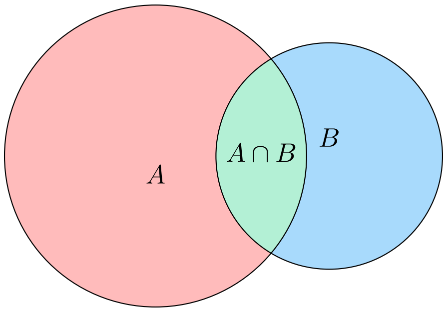
Given the sets $A = \{1, 2, 5, 6, 7\}$ and $B = \{-1, 0, 1, 5, 10, 13, 15\}$, the intersection of A and B is $\{1, 5\}$.

The symbol denoting intersection is $\cap$. An intersection can be formally defined as

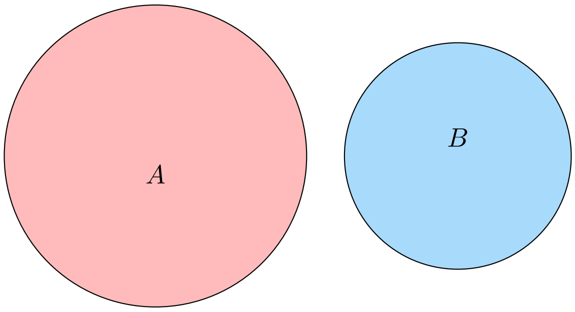
$$A \cap B = \left\{x \mid x \in A \wedge x \in B\right\}$$

(read: ”the intersection of A and B is the set containing all elements x, such that x is in A and x is in B”)

A Venn diagram visualization of $A \cap B$ (green area):



If the intersection of two sets is empty ($A \cap B = \emptyset$), then the sets are said to be *disjoint*:



The *union* of two sets A, B is the set of all elements that are either in A or in B (or both).

Example 14

The union of the sets $A = \{-5, 7, 1\}$ and $B = \{10, -2, -5, 2\}$ is

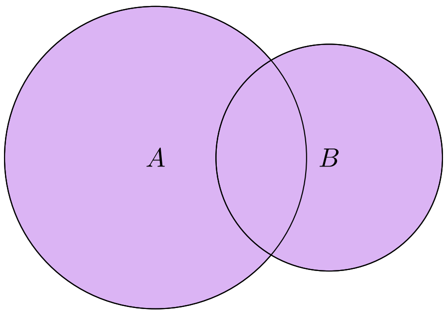
$$A \cup B = \{10, -2, -5, 2, 7, 1\}.$$

The symbol denoting union is $\cup$. A union can be formally defined as

$$A \cup B = \left\{x \mid x \in A \lor x \in B\right\}$$

(read: ”the union of A and B is the set containing all elements x, such that x is in A or x is in B”)

A Venn diagram visualization of $A \cup B$ (purple area):



What is the number of elements in $A \cup B$? Let's look at an example: $A = \{1, 2, 3, 4, 5\}$, $B = \{4, 5, 6, 7\}$. We can see that $|A| = 5$, $|B| = 4$. The union of A, B is $A \cup B = \{1, 2, 3, 4, 5, 6, 7\}$, and $|A \cup B| = 7$. If we just count each element in both sets, we will have 9 elements: $1, 2, 3, 6$ and 7 are each counted once, but 4 and 5 are each counted twice (because they are both in A and in B). To account for this double count, we can simply subtract 2 - the number of elements that are in both sets. This is exactly $|A \cap B|$ (the number of elements in the intersection of A and B). Hence, for two sets A, B, this always holds:

$$|A \cup B| = |A| + |B| - |A \cap B|$$

Note that if A, B are disjoint, $|A \cup B| = |A| + |B|$ (because $|A \cap B| = 0$).

1.2.4 Difference of Sets

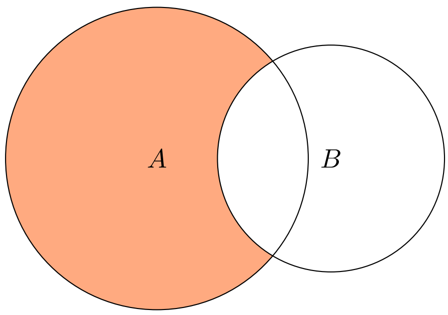
The *difference* of A and B is the set of all elements in A that *are not* elements of B. This is written as $A - B$ (or sometimes $A \setminus B$). Example:

$$A = \{1, 5, 9, 10\}, \ B = \{-3, 2, 5, 9, 13\}, \ A - B = \{1, 10\}$$

Formally:

$$A - B = \left\{x \mid x \in A, \ x \notin B\right\}$$

A Venn diagram visualization of $A - B$ (orange area):



1.2.5 Complement

The *complement* of a set A in a reltaion to a superset $Z \supset A$ is the difference $Z - A$, and is notated A^{c}.

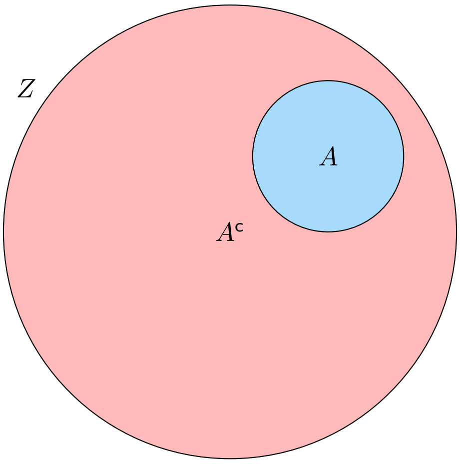
Example 15

$Z = \{1, 2, 3, 4, 5\}, A = \{1, 2, 3\} \Rightarrow A^{\mathsf{c}} = \{4, 5\}$.

Formally:

$$A^{\mathsf{c}} = \left\{x \in Z \mid x \notin A\right\}$$

A Venn diagram representation:



1.2.6 Power Set

The set of all subsets of a given set A is called the *powerset of* A.

Example 16

$A = \{1, 2, 3\}$. All the subsets of A are:

$$\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}.$$

Thus, the power set of A is

$$P(A) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

Note 3

The empty set $\emptyset$ is a subset of all sets. Each set is also a subset of itself.

1.2.7 Important Number Sets

Some important sets of numbers, that will be used frequently in the course (all of them have infinite number of elements):

- **The natural numbers** (symbol: $\mathbb{N}$). These are the numbers $1, 2, 3, \ldots$.
- **The integers** (symbol: $\mathbb{Z}$). These are the "whole numbers" (i.e. not fractions). They include all the natural numbers together with their negatives (i.e. $-1, -2, -3, \ldots$) and 0.
- **The rational** numbers (symbol: $\mathbb{Q}$). As their name suggests, they are ratios between two integers (e.g. $\frac{1}{2}$, $\frac{-5}{3}$, $\frac{7}{13}$).
- **The real numbers** (symbol: $\mathbb{R}$). These are all the numbers on the number line (e.g. $2, \pi, \frac{\sqrt{3}}{17}, \sqrt{5}, -7.2, e^{\pi}$). A proper definition of the real numbers is beyond the scope of this course.
- **The complex numbers** (symbol: $\mathbb{C}$). These are numbers of the form $a + bi$, where a, b are real numbers, and $i = \sqrt{-1}$.

Important Number Sets

Symbol	Name	Definition
$\mathbb{N}$	Natural numbers	$\{1, 2, 3, 4, \dots\}$
$\mathbb{Z}$	Integers	$\{0, \pm x \mid x \in \mathbb{N}\}$
$\mathbb{Q}$	Rational numbers	$\left\{\frac{p}{q} \mid p \in \mathbb{Z}, q \in \mathbb{N}\right\}$
$\mathbb{R}$	Real numbers	
$\mathbb{C}$	Complex numbers	$\left\{a + ib \mid a, b \in \mathbb{R}, i = \sqrt{-1}\right\}$

The relations between these sets are

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$$

(the symbol $\subset$ means "a proper subset")

Note 4

Although each of these sets is infinite, the actual number of elements in $\mathbb{R}$ and $\mathbb{C}$ is bigger than the number of elements in $\mathbb{N}, \mathbb{Z}$ and $\mathbb{Q}$. There are different kinds of infinities.

1.2.8 Cartesian Products

The *cartesian product* of two sets A, B, denoted $A \times B$, is the set of all possible **ordered** pairs, where the first component is an element of A and the second component is an element of B.

Example 17

Consider $A = \{1, 2, 3\}$, $B = \{x, y\}$. Then:

$$A \times B = \left\{(1, x), (1, y), (2, x), (2, y), (3, x), (3, y)\right\}$$

Note 5

Since the elements of a cartesian product are **ordered** pairs, $A \times B \neq B \times A$ (unless $A = B$).

Formally:

$$A \times B = \left\{(a, b) \mid a \in A,\ b \in B\right\}$$

The number of elements in a cartesian product is

$$|A \times B| = |A| \cdot |B|$$

The definition of a cartesian product can be expanded to $n \in \mathbb{N}$ sets $A_1, A_2, \dots, A_n$:

$$A_1 \times A_2 \times \cdots \times A_n = \left\{(a_1, a_2, \dots, a_n) \mid a_1 \in A_1, a_2 \in A_2, \dots, a_n \in A_n\right\}$$

or more compactly:

$$\prod_{i=1}^{n} A_i = \left\{(a_1, a_2, \dots, a_i) \mid a_i \in A_i, i = 1, 2, \dots, n\right\}$$

The symbol $\prod$ is a generalized product notation. It will be discussed in more details later in the course.

A cartesian product of the same set is written in an similar way to a power. For example

$$\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$$

$$\mathbb{R} \times \mathbb{R} \times \mathbb{R} = \mathbb{R}^3$$

These are, respectively, sets of pairs of real numbers, e.g. $(-3, 1), (\pi, 2), (-\frac{\sqrt{7}}{13}, 0)$, and triples of real numbers, e.g. $(1, 2, -\pi), \left(-6, \frac{1}{\sqrt{\pi}}, 0.2\right), \left(\frac{1}{51}, \sqrt{3}, -4\right)$.

Example 18

$$A = \{a, b\} \;\Rightarrow\; A^3 = \left\{(a,a,a), (a,a,b), (a,b,a), (a,b,b), (b,a,a), (b,a,b), (b,b,a), (b,b,b)\right\}.$$
$$B = \{1, 2, 3\} \;\Rightarrow\; B^2 = \left\{(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3)\right\}.$$

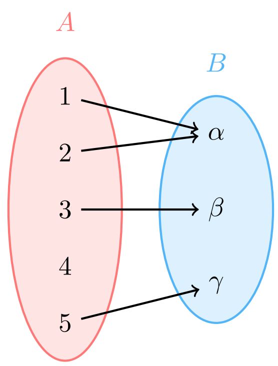
1.3 Functions

1.3.1 Relations

A *relation* between two sets A, B is a way to "connect" the elements in the two sets in pairs. It is a subset of the cartesian product $A \times B$.

Example 19

A relation R between the sets $A = \{1, 2, 3, 4, 5\}$ and $B = \{\alpha, \beta, \gamma\}$.

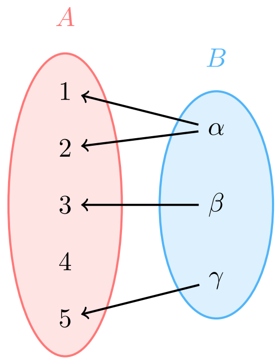


Notice how not all elements are connected (for example, 4 is not connected to any element in B), and some elements in A are connected to the same element in B ($1, 2$ are both connected to α). The pairs making up the relation R are $(1, \alpha), (2, \alpha), (3, \beta), (5, \gamma)$, which are all elements of the cartesian product $A \times B$.

Reversing the connections yields R^{-1}, which is a subset of $B \times A$.

Example 20

The converse relation R^{-1} for the above relation R between A and B.



The pairs making up the relation R^{-1} are $(\alpha, 1), (\alpha, 2), (\beta, 3), (\gamma, 5)$, which are all elements of the cartesian product $B \times A$.

1.3.2 Basic Definition of a Function

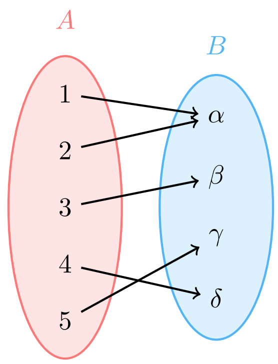
A *function* between the sets A, B is a relation in which for every element $a \in A$ there is exactly **one** connection to an element $b \in B$.

Note 6

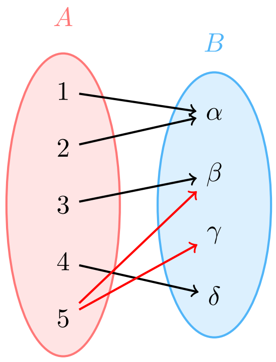
A function can have more than one element $a \in A$ connected to the same element $b \in B$. The only restriction is that no element $a \in A$ is connected to **more than one** element $b \in B$.

Example 21

A valid function from A to B:



An invalid function from A to B (connections in red cause invalidity):



1.3.3 Domain and Image of a Function

We notate that a function f connects each element $a \in A$ to an element $b \in B$ by writing

$$f : A \to B.$$

The set A is then the *domain* of f, while B is the *image* of f.

Another notation is writing the function as acting on an *argument*:

$$f(a) = b.$$

(read: "f of a is equal to b")

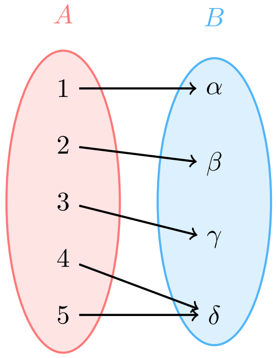
WRITE: Real functions, plotting functions.

1.3.4 Injective, Surjective and Bijective Functions

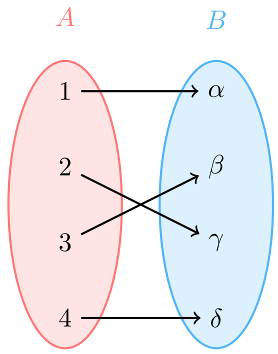
A function is called *injective* if each of the elements in its **image** is connected to by a single element in its **domain**.

Example 22

A function that is **not injective**:



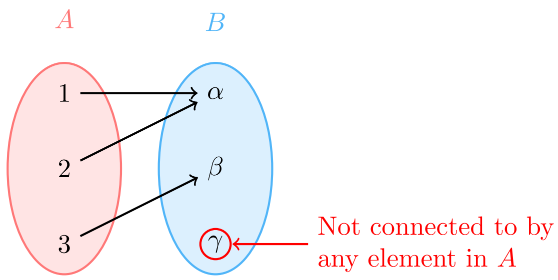
An injective function:



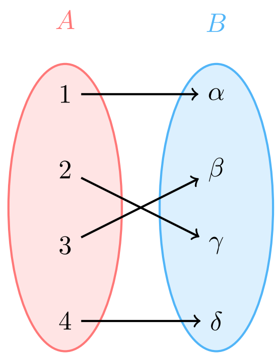
A function is *surjective* if each of the elements in its image is connected to by **at least** one element from its domain.

Example 23

A function that is **not surjective**:



A sujective function:

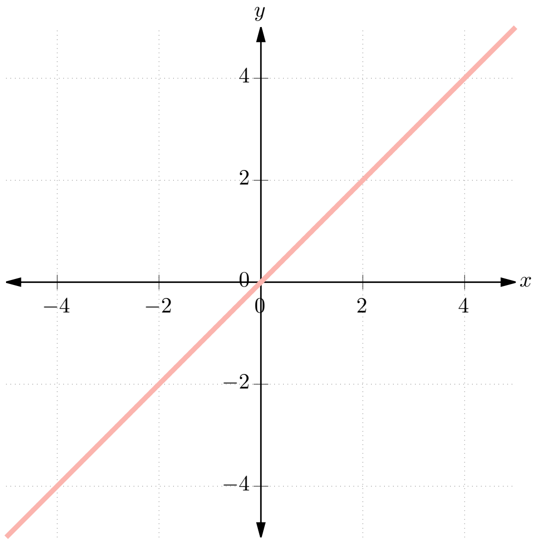


A *bijective* function is a function that is **both** injective and surjective. The inverse of a bijective function is also a function.

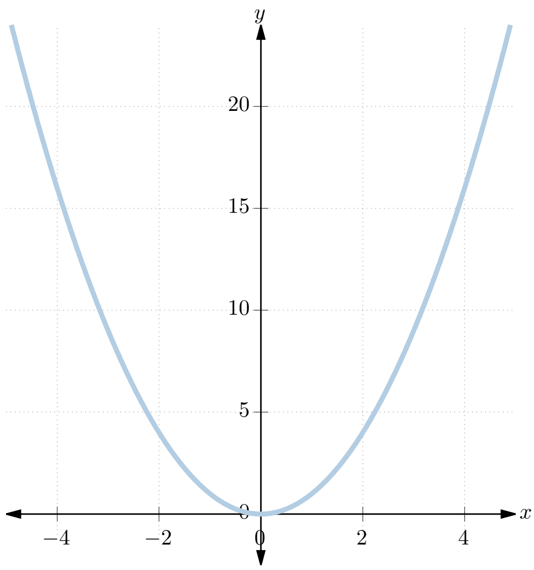
Example 24

A few functions $f : \mathbb{R} \to \mathbb{R}$ that are either injective, surjective, bijective or none of the above:

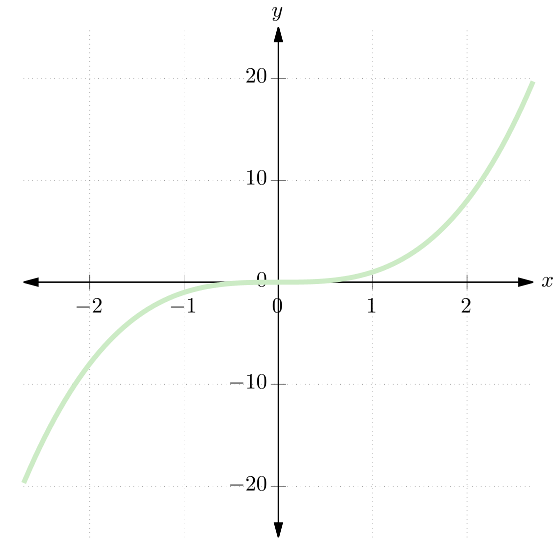
- $f(x) = x$, injective and surjective $\Rightarrow$ bijective.



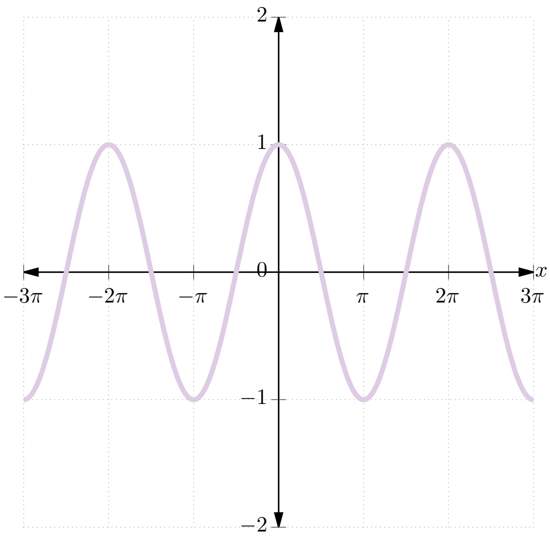
- $f(x) = x^2$, not injective nor surjective.
  Not injective: For example, $f(3) = 3^2 = 9 = (3)^2 = f(-3)$. In fact, each number $0 \neq x \in \mathbb{R}$ has a non-unique image, since $x^2 = (-x)^2$.
  Not surjective: Over $\mathbb{R}$, any negative number is not connected to by any input in the domain.



- $f(x) = x^3$: injective and surjective $\Rightarrow$ bijective.



- $f(x) = \cos(x)$, not injective nor subjective.
  Not injective: For example, $1 = f(0) = f(2\pi)$.
  Not surjective: The actual image of $\cos(x)$ is $[-1, 1]$, and over this interval it is surjective.



1.3.5 Multivariable Functions

Functions may have several arguments and return several arguments.

Example 25

The follows functions take as input three real numbers, and return a single real number ($f : \mathbb{R}^3 \to \mathbb{R}$). The return value of some functions for a triple of real numbers, $(-5, 7, 1)$, are:

- $f(x, y, z) = x + y + z \Rightarrow f(-5, 7, 1) = -5 + 7 + 1 = 3$
- $f(x, y, z) = x^2 - y^2 \Rightarrow f(-5, 7, 1) = 25 - 49 = -24$

- $f(x, y, z) = \frac{x}{\sqrt{y}+z} \Rightarrow f(-5, 7, 1) = \frac{5}{\sqrt{7}+1}$

Example 26

$f : \mathbb{Z} \times \mathbb{N} \to \mathbb{Q}$ is defined as follows: $f(p, q) = \frac{p}{q}$. The return value of f for the some example inputs:

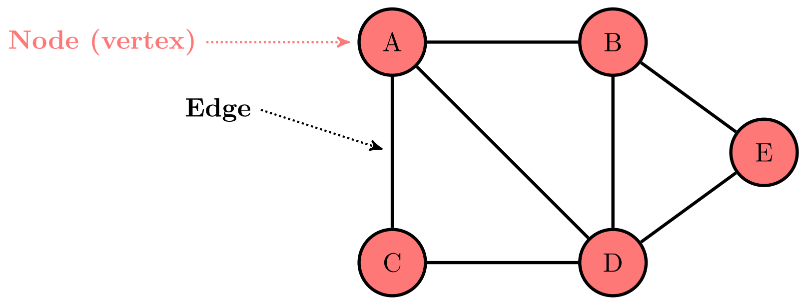
- $(1, 2) \Rightarrow \frac{1}{2} = 0.5$
- $(-5, 2) \Rightarrow \frac{-5}{2} = -2.5$
- $(0, 13) \Rightarrow \frac{0}{13} = 0$

1.4 Graphs

Graphs are mathematical structures composed of *nodes* (or *vertices*, singular *vertex*) connected to eachother by *edges*.

Example 27

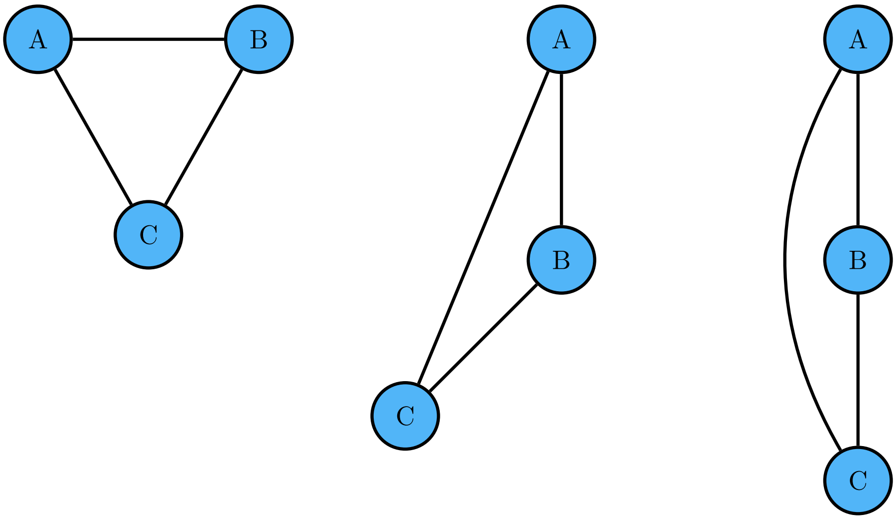
A graph with 5 nodes and 7 edges:



In the graphical representation of a graph, the actual position of nodes does not matter - what matters are the connections (edges) between them.

Example 28

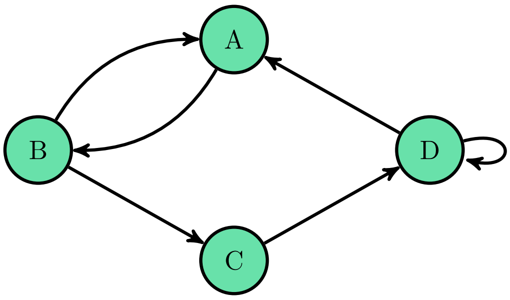
The follows are different representations of **the same** graph (3 nodes and 3 edges):



A graph where all edges have some direction is called a *directed graph*.

Example 29

A directed graph with 4 nodes and 6 edges:



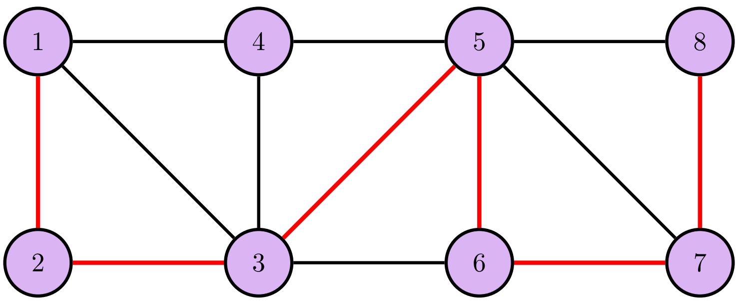
Note 7

In a directed graph, an edge can "loop" back to a node. See node D in the example above.

A *path* in a graph is a sequence of edges in which each edge shares a vertex with the previous edge (except the first edge).

Example 30

In the following graph the edges highlighted in red form a path:

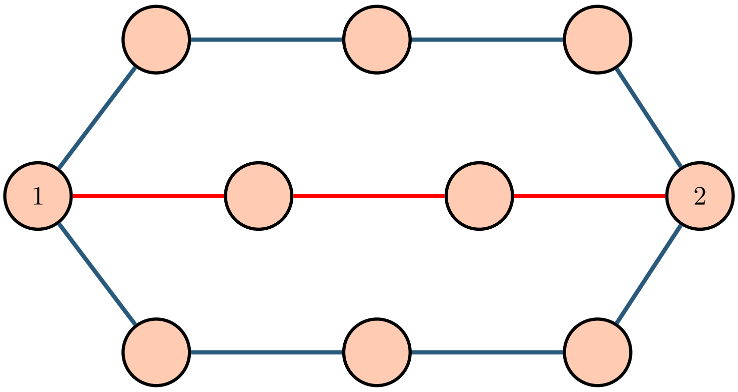


When the start and end vertices coincide the path is known as a *circle*. A directed circle is known as a *cycle*.

If one or more pathes exist between two vertices a, b in a graph, the number of edges in the shortest path is defined to be the *distance* between the two vertices, and is denoted as $\text{dist}(a, b)$.

Example 31

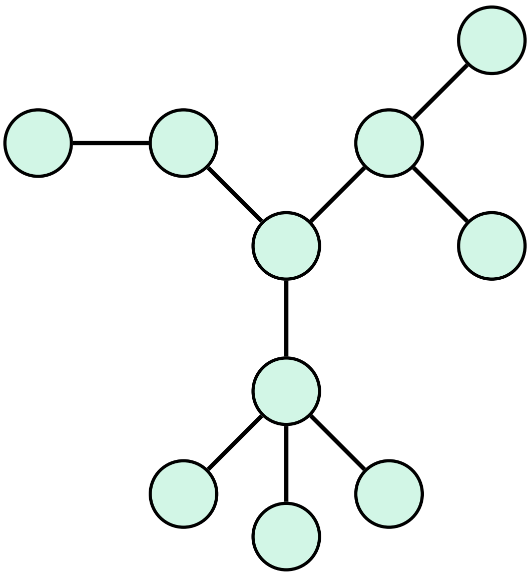
In the following graph three paths between vertices 1 and 2 are shown. The number of edges in the shortest path, highlighted in red, is defined as the distance $\text{dist}(1, 2)$, and is equal to 3.



A *tree* is a graph with no circles.

Example 32

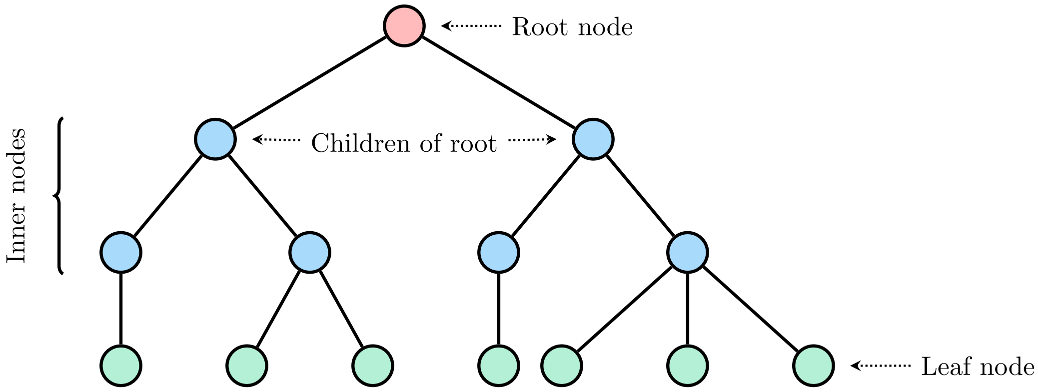
A tree (notice that no circles are present):



Some trees have a distinctive *root* node, and are known as *rooted trees*. A node that is "branched" from a higher level node is called a *child node*. The last level nodes are called *leaves* (singular: leaf). The rest of the nodes are known as *inner nodes*.

Example 33

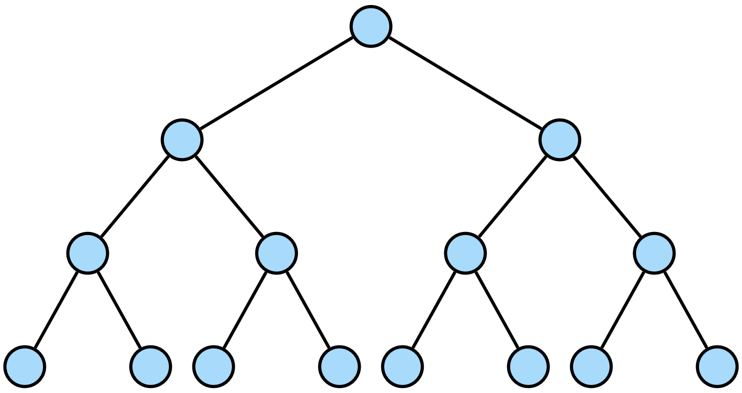
A rooted tree, with the root node highlighted in red and the leaves in green:



Trees with 2 children per node are known as *binary trees*. Similarily, trees can be ternary, quaternary, etc.

Example 34

A binary tree:

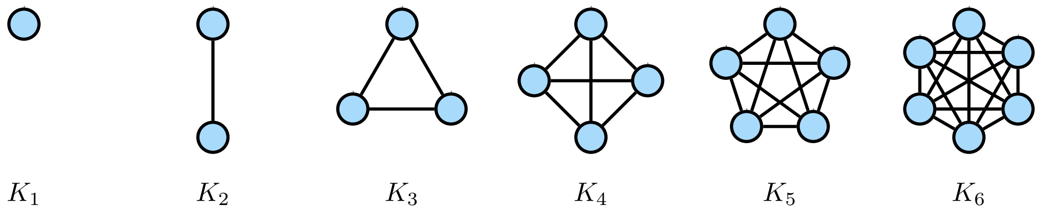


Rooted trees are used to describe hierarchies, e.g. in biological systematics, organisations or nested directories of data.

The *complete graph* K_n is the graph with n vertices where every pair of different vertices is connected by an edge (Also called a *clique*).

Example 35

The cliques $K_1, \ldots, K_6$:



Vectors 2

2.1 Introduction

There are three common ways to describe a vector:

- **"The Physicist's definition"**: An object that has a magnitude and a direction (a geometric definition).
- **"The Computer scientists' definition"**: An array of numbers (an arithmetic definition).
- **"The Mathematician's definition"**: An element of a vector space (an abstract definition).

In this course we will introduce the physicist's (geometric) approach first, then the computer scientists' (arithmetic) approach, linking the two approaches together.
The mathematician's (abstract) approach will remain mostly untouched until the last chapter where we generalize the definition of a vector.

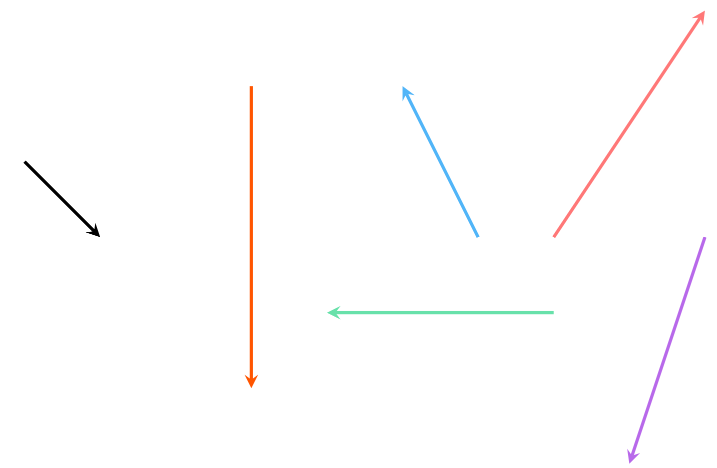
2.2 Geometric Approach

2.2.1 Basics

A *vector* is an object that has a *magnitude* (also a *norm*, or a *length*), and a *direction* (also an *argument*). A vector starts at a point (called its *origin*), and ends at a point. We draw vectors as arrows starting at the start point with their head at the end point.

Example 36

Some 2-dimensional vectors having different starting and ending points:



In physics (and this course), vectors are usually notated by undercase Latin letters with a small right-facing arrow drawn above them, i.e.:

$$\vec{u}, \quad \vec{v}, \quad \vec{x}, \quad \vec{a}, \quad \dots$$

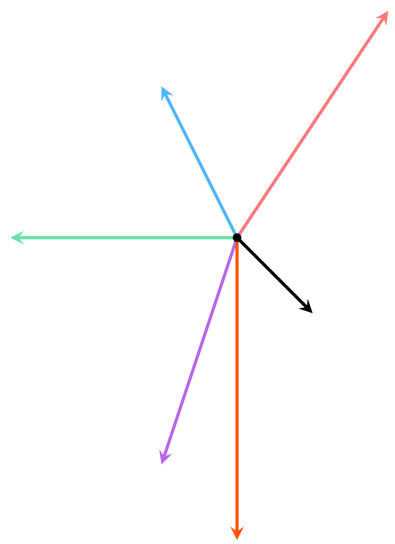
However, mathematicians tend to write vectors as undercase Latin letters that are either bold faced, with or without an underline, i.e.:

$$\boldsymbol{u}, \quad \boldsymbol{v}, \quad \boldsymbol{x}, \quad \boldsymbol{a}, \quad \dots$$
$$\underline{\boldsymbol{u}}, \quad \underline{\boldsymbol{v}}, \quad \underline{\boldsymbol{x}}, \quad \underline{\boldsymbol{a}}, \quad \dots$$

We consider all vectors to originate in the same point (which we call the *origin*).

Example 37

The 2-dimensional vectors from before, placed such that they originate from the same point:

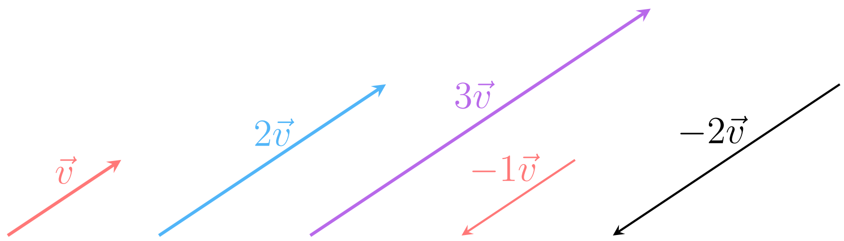


2.2.2 Scaling a Vector, Adding Vectors

Scaling a vector changes its magnitude without changing its direction. This is done by multiplying the vector by a real number, which in this context we call a *scalar*.

Example 38

Scaling a vector $\vec{v}$ by $2, 3, -1$ and -2:



Note 8

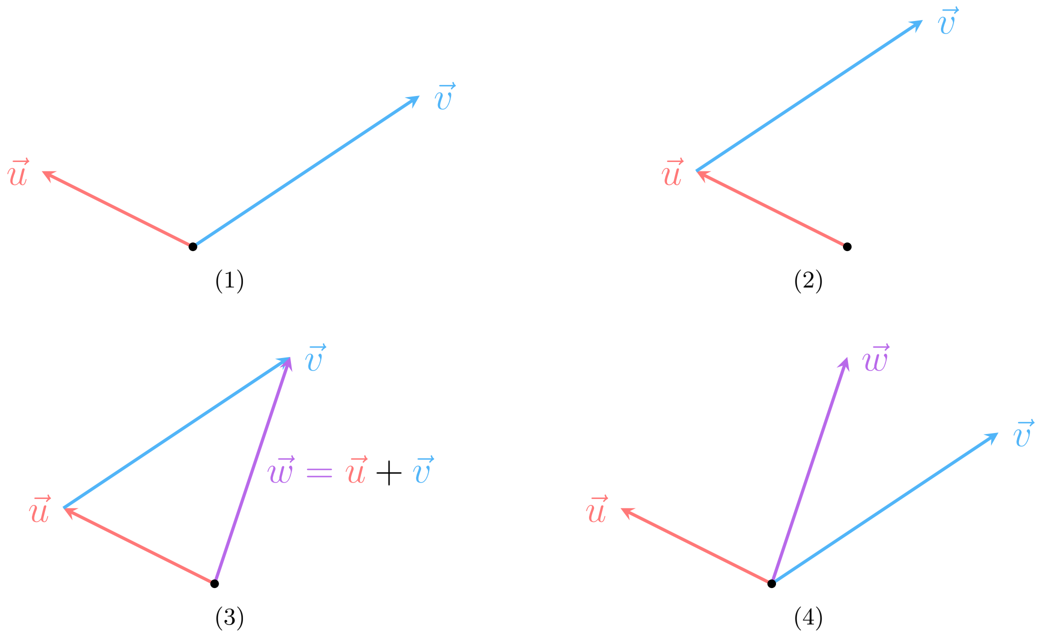
Notice that scaling a vector by a negative number reverses its direction.

Adding two vectors is done by placing the origin of one vector at the head of the other vector. The addition result is a vector starting at the first vector's origin and ending at the second vector's head.

Example 39

Adding two vectors $\vec{u}$ and $\vec{v}$ according to the above scheme:

1. The two vectors $\vec{u}$ and $\vec{v}$ originating from the same point
2. Moving $\vec{v}$ so that its origin is placed at the head of $\vec{v}$.
3. Drawing a vector $\vec{w} = \vec{u} + \vec{v}$ from the origin of $\vec{u}$ to the head of $\vec{v}$.
4. Moving $\vec{v}$ back to the origin.



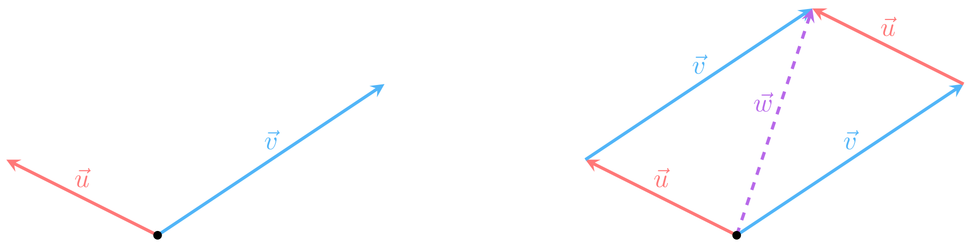
The order of addition does not change the result:

$$\vec{u} + \vec{v} = \vec{v} + \vec{u}.$$

This property is known as *commutativity*.

Example 40

Adding the vectors $\vec{u}, \vec{v}$ in two ways: first by moving $\vec{v}$ so that its origin is at the head of $\vec{u}$, and then by moving $\vec{v}$ such that its origin is at the head of $\vec{u}$. Both additions result in the same vector $\vec{w}$. The drawn addition method is known as the *parallelogram* method.



(1) (2)

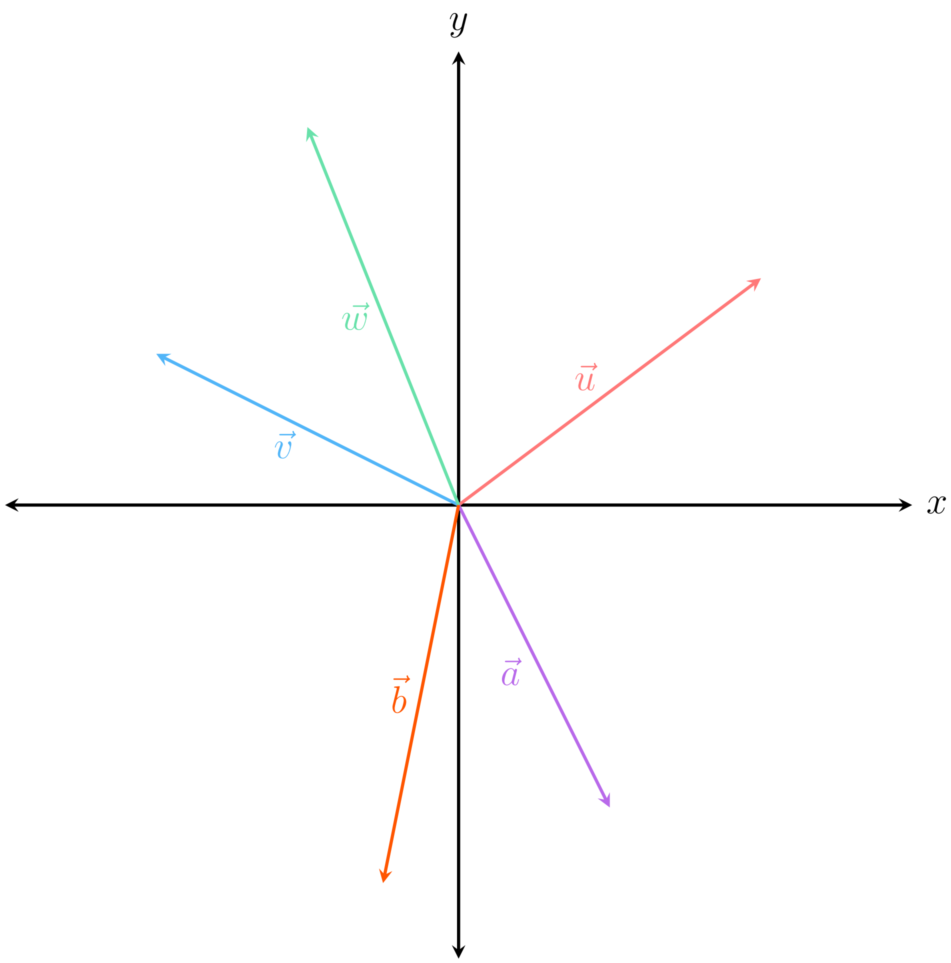
One special vector is the *zero vector* $\vec{0}$, which has a norm of 0 and no direction. The zero vector is neutral to addition:

$$\vec{v} + \vec{0} = \vec{0} + \vec{v} = \vec{v}.$$

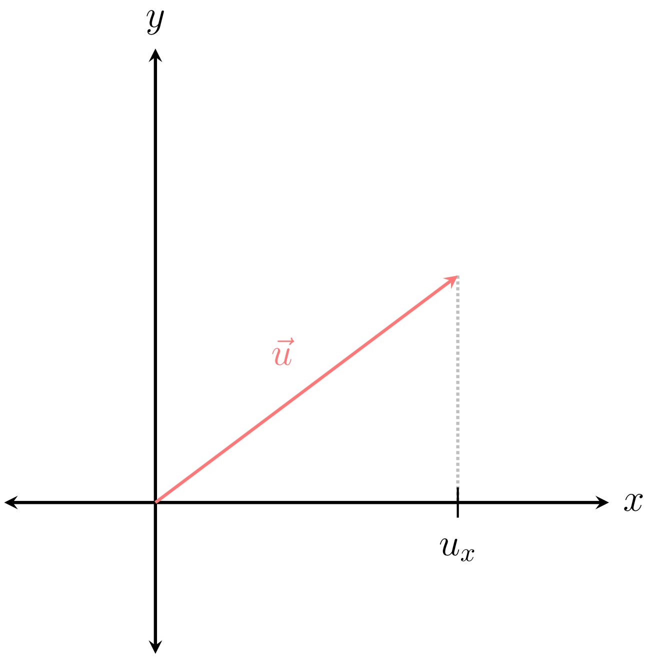
2.3 Algebraic Approach

2.3.1 Basics

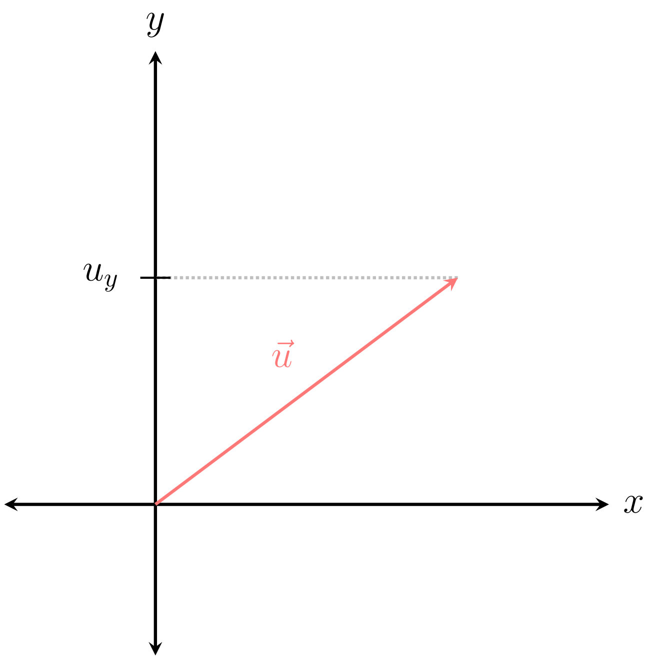
We can place all vectors on an axis system, such that their origins are placed at $x = 0, y = 0$:



For each vector, we can drop a line perpendicular from its head to the x-axis. We label the intercept point with the x-axis as u_x:



Similarly, we drop a line from the head of the vector perpendicular to the y-axis, and label the intercept point as u_y:



We write the vector as a *column vector* as follows:

$$\vec{u} = \begin{pmatrix} u_x \\ u_y \end{pmatrix}$$

we refer to the entries of the vector as its *components* or *coordinates*. The vector $\vec{u}$ is an element in the set $\mathbb{R}^2$.

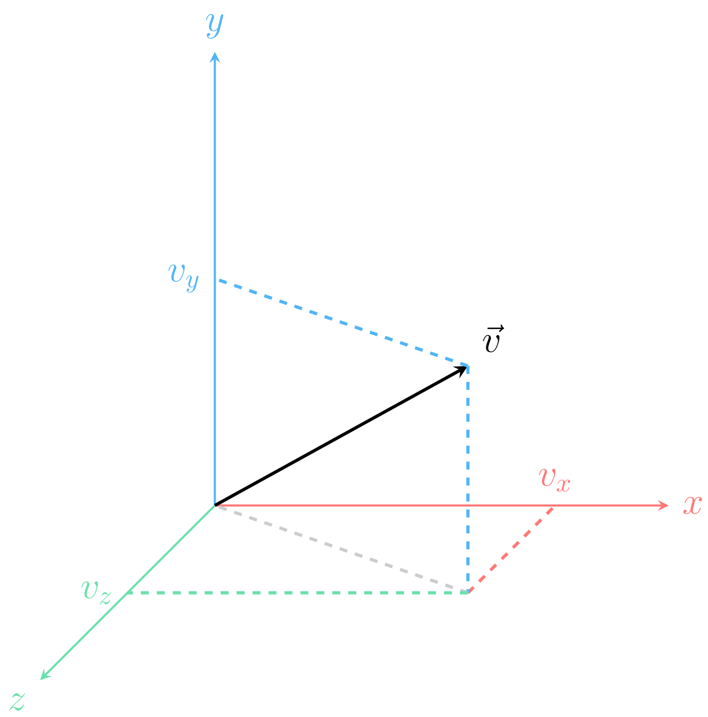
Note 9

We can also write the vector as a *row vector*:

$$\vec{u} = (u_x, u_y)$$

but in some applications (e.g. Tensor analysis) this has a different meaning, and so in this course we will stick with column vectors.

In 3-dimensions (also called $\mathbb{R}^3$) the principal is the same:



As a column vector, the zero vector has all-zero components. In $\mathbb{R}^2$ this means:

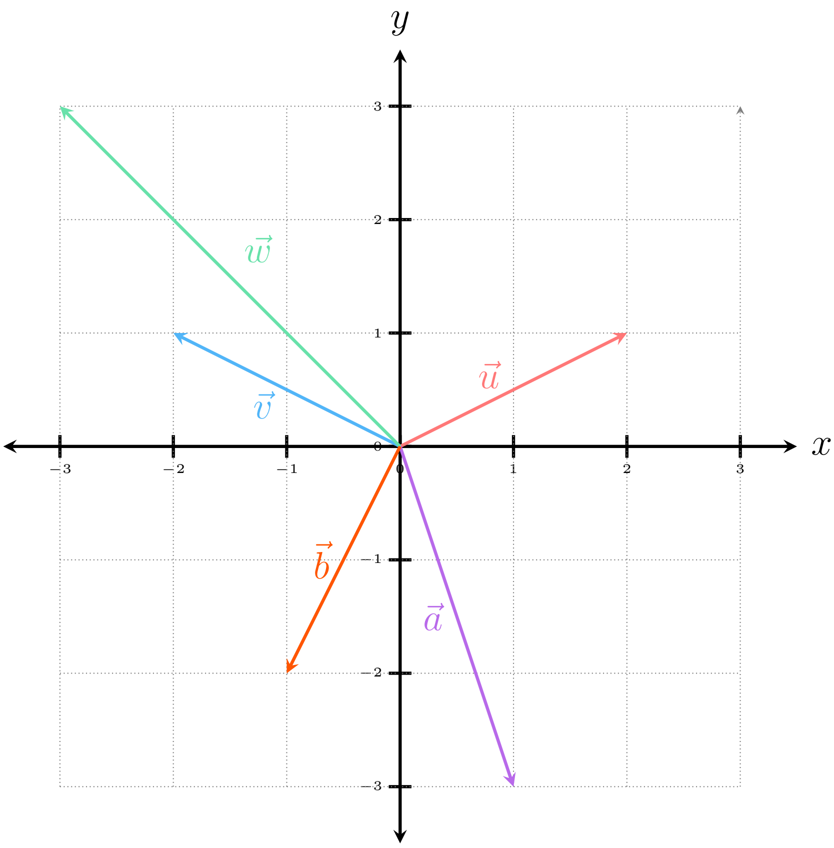
$$\vec{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and in $\mathbb{R}^3$:

$$\vec{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Example 41

For each of the follows vectors in $\mathbb{R}^2$, we will find its components:



$$\vec{u} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \vec{v} = \begin{pmatrix} -2 \\ 1 \end{pmatrix} \quad \vec{w} = \begin{pmatrix} -3 \\ 3 \end{pmatrix} \quad \vec{a} = \begin{pmatrix} 1 \\ -3 \end{pmatrix} \quad \vec{b} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}$$

2.3.2 Scaling a Vector, Adding Vectors

Column vectors are scaled by multiplying each of their components by the same scalar.

Example 42

$$\vec{u} = \begin{pmatrix} 1 \\ -3 \end{pmatrix} \quad \Rightarrow \quad 2\vec{u} = \begin{pmatrix} 2 \\ -6 \end{pmatrix}, \quad -3\vec{u} = \begin{pmatrix} -3 \\ 9 \end{pmatrix}$$

$$\vec{v} = \begin{pmatrix} -2 \\ 0.5 \\ 1 \end{pmatrix} \quad \Rightarrow \quad 4\vec{v} = \begin{pmatrix} -8 \\ 2 \\ 4 \end{pmatrix}, \quad -7\vec{u} = \begin{pmatrix} 14 \\ -3.5 \\ -7 \end{pmatrix}$$

Two column vectors are added by adding their components *element-wise*.

Example 43

$$\vec{u} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \vec{v} = \begin{pmatrix} -1 \\ 4 \end{pmatrix}, \quad \vec{w} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\vec{a} = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} -1 \\ 4 \\ 0 \end{pmatrix}$$

$$\Downarrow$$

$$\vec{u} + \vec{v} = \begin{pmatrix} 0 \\ 6 \end{pmatrix}$$

$$\vec{u} + \vec{w} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

$$\vec{v} + \vec{w} = \begin{pmatrix} -1 \\ 5 \end{pmatrix}$$

$$\vec{a} + \vec{b} = \begin{pmatrix} 0 \\ 6 \\ -3 \end{pmatrix}$$

Note 10

Subtracting a vector from another vector is like adding the first vector to the inverse of the second vector, i.e.

$$\vec{u} - \vec{v} = \vec{u} + (-\vec{v}),$$

where $-\vec{v}$ is simply $-1 \cdot \vec{v}$.

Of course, everything written so far is true for vectors of any natural number of components ($n = 1, 2, 3, \dots$). The number of components in a vector is called its *dimension*.

Example 44

Addition of vectors in $\mathbb{R}^4$:

$$\begin{pmatrix} 1 \\ -2 \\ 0 \\ 7 \end{pmatrix} + \begin{pmatrix} 0 \\ 5 \\ -2 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ -2 \\ 9 \end{pmatrix}$$

Addition of vectors in $\mathbb{R}^5$:

$$\begin{pmatrix} 1 \\ -1 \\ 2 \\ 0 \\ -1 \end{pmatrix} + \begin{pmatrix} -2 \\ 3 \\ 0 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ 2 \\ -1 \\ 0 \end{pmatrix}$$

General Addition of two vectors in $\mathbb{R}^n$:

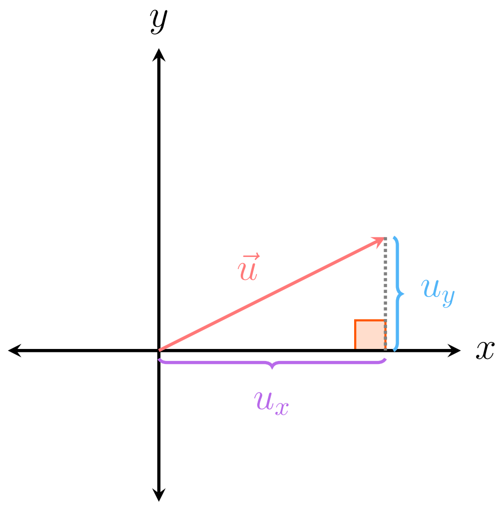
$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

Note 11

Addition of vectors of different dimensions is **undefined**!

2.4 Norm (Length) and Argument (Angle) of a Vector

The *norm* (or *length*) of a 2-dimensional vector $\vec{u}$, denoted as $\|\vec{u}\|$, can be calculated via Pytharogas' theorem:



This results in the norm being:

$$\|\vec{u}\| = \sqrt{u_x^2 + u_y^2}.$$

Example 45

The norm of $\vec{u} = \begin{pmatrix} 3 \\ -4 \end{pmatrix}$ is

$$\begin{aligned} \|\vec{u}\| &= \sqrt{3^2 + (-4)^2} \\ &= \sqrt{9 + 16} \\ &= \sqrt{25} \\ &= 5. \end{aligned}$$

This definition can be expanded to $\mathbb{R}^3$:

$$\|\vec{v}\| = \sqrt{v_x^2 + v_y^2 + v_z^2}.$$

Challange 2

Show that Pytharogas' theorem as written here does indeed hold for 3-dimensional distance of two dots (hint: use a 3-dimensional box).

A generalization of this definition for any vector in $\mathbb{R}^n$ is:

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} = \sqrt{\sum_{i=1}^{n} v_i^2}.$$

A *normalized vector* is a vector with norm=1. To normalize any vector $\vec{v}$ in $\mathbb{R}^n$ (i.e. get a vector pointing in the same direction as $\vec{v}$ but with norm=1), we simply divide the vector $\vec{v}$ by its norm. We denote such vector by $\hat{v}$:

$$\hat{v} = \frac{1}{\|\vec{v}\|}\vec{v}.$$

Example 46

Let's normalize the vector $\vec{v} = \begin{pmatrix} -1 \\ 0 \\ 5 \\ -3 \end{pmatrix}$. First, we find its norm:

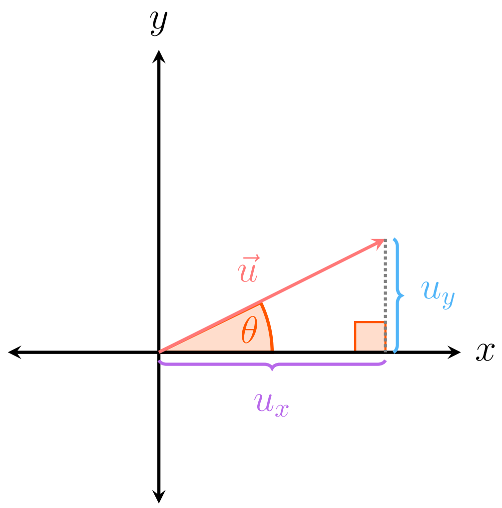
$$\begin{aligned} \|\vec{v}\| &= \sqrt{(-1)^2 + 0^2 + 5^2 + (-3)^2} \\ &= \sqrt{1 + 0 + 25 + 9} \\ &= \sqrt{35}. \end{aligned}$$

Therefore,

$$\|\vec{v}\| = \frac{1}{\sqrt{35}} \begin{pmatrix} -1 \\ 0 \\ 5 \\ -3 \end{pmatrix}.$$

The *argument* (or *angle*) of a vector in $\mathbb{R}^2$ in relation to the x-axis (counter-clockwise) can be found using trigonometry: in the example $\vec{u}$ above, the angle θ opposing u_y is the angle we're after, and its tangent is

$$\tan(\theta) = \frac{u_y}{u_x}.$$



The norm and argument of a vector are called its *polar coordinates* (as opposed to its *cartesian coordinates*, which are its (x, y)-components).

Usually in the context of polar coordinates, the norm is denoted as r or R.

Transforming a vector $\vec{u}$ from its polar coordinates (r, θ) to its cartesian coordinates is done as follows:

$$\begin{aligned} u_x &= r \cdot \cos(\theta), \\ u_y &= r \cdot \sin(\theta). \end{aligned}$$

The inverse transformation is (as seen above):

$$r = \sqrt{u_x^2 + u_y^2},$$

$$\theta = \arctan\left(\frac{u_y}{u_x}\right),$$

where arctan is the inverse of $\tan(\theta)$, i.e. if $x = \tan(\theta)$, then $\theta = \arctan(x)$ (sometimes this function is called $\tan^{-1}$).

Example 47

The vector $\begin{pmatrix} 3 \\ -2 \end{pmatrix}$ has the follows polar coordinates:

$$\begin{aligned} r &= \sqrt{3^2 + (-2)^2} = \sqrt{9+4} = \sqrt{13}, \\ \theta &= \arctan\left(\frac{-2}{3}\right) \approx 5.7\ [\text{rad}] = 326.31^\circ. \end{aligned}$$

The vector with polar coordinates $\left(2, \frac{\pi}{2}\right)$ (the argument is given in radians), has the follows cartesian coordinates:

$$\begin{aligned} u_x &= 2 \cdot \cos\left(\frac{\pi}{2}\right) = 0, \\ u_x &= 2 \cdot \sin\left(\frac{\pi}{2}\right) = 2. \end{aligned}$$

(remember that $\cos\left(\frac{\pi}{2}\right) = 0,\ \sin\left(\frac{\pi}{2}\right) = 1$)

2.5 Spaces and Subspaces

2.5.1 Linear Dependency and Linear Combinations

Two vectors that point in the same direction are said to be *linearly dependent*. Pointing in the same direction means that the vectors are scales of each other:

$$\vec{u} = \alpha\vec{v}, \qquad \alpha \in \mathbb{R}.$$

Example 48

The follows vector pairs are all linearly **dependent**:

$$\begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 6 \end{pmatrix} \qquad \begin{pmatrix} 3 \\ -2 \end{pmatrix}, \begin{pmatrix} -3 \\ 2 \end{pmatrix} \qquad \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ 4 \end{pmatrix} \qquad \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} -3 \\ -6 \\ -9 \end{pmatrix}$$

$$\begin{pmatrix} 0 \\ 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 0 \\ 4 \end{pmatrix} \qquad \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \\ 6 \\ 8 \\ 10 \end{pmatrix}$$

A *linear combination* of two vectors is the sum of the vectors each scaled by some scalar:

$$\vec{w} = \alpha\vec{u} + \beta\vec{v}$$

here, the vector $\vec{w}$ is a linear combination of the vectors $\vec{u}$ and $\vec{v}$, which are scaled by the scalars α and β, respectively.

Example 49

Let's take the two vectors $\vec{u} = \begin{pmatrix} 1 \\ 0 \\ -3 \end{pmatrix}$ and $\vec{v} = \begin{pmatrix} -2 \\ 1 \\ -5 \end{pmatrix}$. Together with the scalars $\alpha = \frac{1}{2}$ and $\beta = -3$, a linear combination of $\vec{u}$ and $\vec{v}$ can be constructed:

$$\begin{aligned} \vec{w} &= \alpha\vec{u} + \beta\vec{v} \\ &= \frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ -3 \end{pmatrix} - 3\begin{pmatrix} -2 \\ 1 \\ -5 \end{pmatrix} \\ &= \begin{pmatrix} 0.5 \\ 0 \\ -1.5 \end{pmatrix} - \begin{pmatrix} 6 \\ -3 \\ 15 \end{pmatrix} \\ &= \begin{pmatrix} 6.5 \\ -3 \\ 13.5 \end{pmatrix} \end{aligned}$$

For any number $n \in \mathbb{N}$ of vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ and scalars $\alpha_1, \alpha_2, \dots, \alpha_n$, the resulting linear combination is simply

$$\begin{aligned} \vec{w} &= \alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n \\ &= \sum_{i=1}^{n} \alpha_i\vec{v}_i. \end{aligned}$$

Note 12

Some of the scalars α_i can be zero. If all of them are zero, then the resulting sum is necessarily the zero vector $\vec{0}$. This is, however, a very boring case.

A set of $n \in \mathbb{N}$ vectors $\{\vec{v}_i\}$ are said to be linearly independent if there exist a set of n scalars $\{\alpha_i\}$, **not all of them zeros**, such that

$$\sum_{i=1}^{n} \alpha_i\vec{v}_i = \vec{0}.$$

Example 50

Let's see how this definition works for two general 2-dimensional vectors $\vec{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$: if for some scalars α, β (not both of them zero!) the linear combination $\alpha\vec{u} + \beta\vec{v} = \vec{0}$.

There are three cases: either $\alpha = 0, \beta = 0$ or both α and β are nonzero.

1. The first case means that
$$\beta\vec{v} = \vec{0} \Rightarrow \vec{v} = \vec{0},$$
and thus $\vec{v}$ is the zero vector, which is always a scale of any vector (with a zero scalar).
2. In the case $\beta = 0$ we get, similarily to the previous case, that $\vec{u}$ is the zero vector.

3. When both α and β are non-zero,

$$\alpha\vec{u} = \vec{0} - \beta\vec{v} = -\beta\vec{v},$$

which in turn means

$$\vec{u} = -\frac{\beta}{\alpha}\vec{v}.$$

But this is exactly the definition of linear dependency for two vectors: they are simply a scale of each other!

For a set of N-dimensional vectors, any linear combination of $M > N$ vectors will necessarily be linearly depended.

Example 51

For 2-dimensional vectors, every set of 3 or more vectors is linearly dependent.

Let's look at a specific example: $\vec{u} = \begin{pmatrix}1\\2\end{pmatrix}, \vec{v} = \begin{pmatrix}-2\\0\end{pmatrix}, \vec{w} = \begin{pmatrix}0\\1\end{pmatrix}$.

For the scalars $\alpha = 2, \beta = 1, \gamma = -4$, we get

$$\begin{aligned}\alpha\vec{u} + \beta\vec{v} + \gamma\vec{w} &= 2\begin{pmatrix}1\\2\end{pmatrix} + \begin{pmatrix}-2\\0\end{pmatrix} - 4\begin{pmatrix}0\\1\end{pmatrix}\\ &= \begin{pmatrix}2\\4\end{pmatrix} + \begin{pmatrix}-2\\0\end{pmatrix} + \begin{pmatrix}0\\-4\end{pmatrix}\\ &= \begin{pmatrix}2-2+0\\4+0-4\end{pmatrix}\\ &= \begin{pmatrix}0\\0\end{pmatrix}\\ &= \vec{0},\end{aligned}$$

which means that these three vectors are linearly dependent.

Note 13

The condition for linear dependency of a set of vector can be reformulated as follows: if any of the vectors can be written as a linear combination of all the other vectors, then the set is linearly dependent (and any vector in the set could be written as a linear combination of all the other vectors).

Challange 3

Show that if a set of vectors (of any dimension and number of vectors) is linearly dependent, any vector in the set could be written as a linear combination of the rest of the vectors.

2.5.2 Vector Spaces and Bases

As we saw in the previous subsection, if two 2-dimensional vectors are linearly independent, then any other 2-dimensional vector can be written as a linear combination of these two vectors, i.e.

$$\vec{w} = \alpha\vec{u} + \beta\vec{v}.$$

Since any vector in $\mathbb{R}^2$ can be written as a linear combination of two linearly independent vectors in $\mathbb{R}^2$, we say that such vectors **span** $\mathbb{R}^2$.

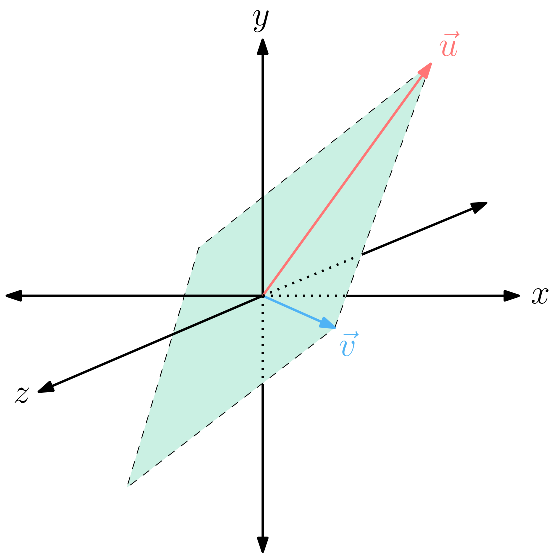
We already implicitly use two such vectors: $\begin{pmatrix}1\\0\end{pmatrix}$ and $\begin{pmatrix}0\\1\end{pmatrix}$ are linearly independent, and any 2-dimensional vector $\vec{v} = \begin{pmatrix}x\\y\end{pmatrix}$ can be written as a linear combination of these two vectors, with scalars $\alpha = x, \beta = y$:

$$\begin{pmatrix}x\\y\end{pmatrix} = x\begin{pmatrix}1\\0\end{pmatrix} + y\begin{pmatrix}0\\1\end{pmatrix}.$$

This idea can be expanded to 3-dimensional vectors: the entire space $\mathbb{R}^3$ can be spanned by any set of three vectors in $\mathbb{R}^3$ which are linearly independent. And of course, this applies to any n-dimensional space $\mathbb{R}^n$: any set of n linearly independent vectors in $\mathbb{R}^n$ completely span $\mathbb{R}^n$.

What about, for instance, two linearly independent 3-dimensional vectors? They can't span the entire 3-dimensional space, but they do span an infinite plane inside 3-dimensional space **that goes through the origin**. Such a plane is called a *subspace* of the 3-dimensional space.

In the figure below the two vectors $\vec{u}$ and $\vec{v}$, both in $\mathbb{R}^3$ span a plane (in green) that goes through the origin $\vec{0} = \begin{pmatrix}0\\0\\0\end{pmatrix}$:



(Figure recreated from `https://www.ck12.org/book/ck-12-math-analysis/section/5.4/`)

Similarly, in n-dimensional space ($\mathbb{R}^n$), any $m < n$ linearly independent vectors span a space $\mathbb{R}^m$ that is a subspace of $\mathbb{R}^n$, that goes through the origin.

Note 14

Any subspace goes through the origin of the original space!

Challange 4

What is the geometric shape of a subspace of $\mathbb{R}^3$ that is spanned by a single vector?

A set of linearly independent vectors that span a space is called a *basis* of that space.

Example 52

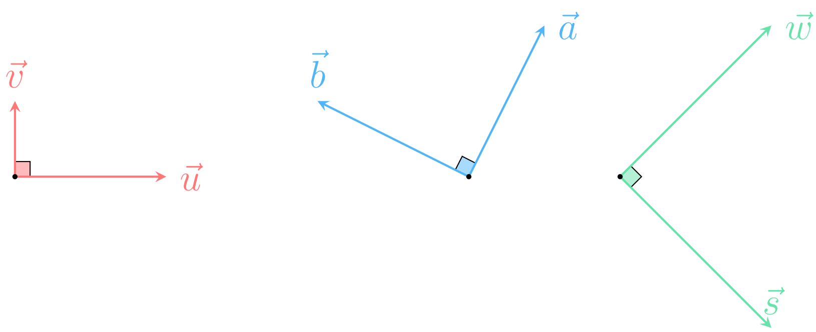
The vectors $\begin{pmatrix}1\\-2\end{pmatrix}$ and $\begin{pmatrix}0\\\frac{1}{2}\end{pmatrix}$ are a basis of $\mathbb{R}^2$.

The vectors $\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}-\frac{1}{4}\\0\\2\end{pmatrix}$ and $\begin{pmatrix}1\\-1\\2\end{pmatrix}$ are a basis of $\mathbb{R}^3$.

If all the basis vectors are *orthogonal* to each other[1], then the basis is said to be an *orthogonal basis*.

Example 53

In the following figure, each of the sets of vectors $\{\vec{u}, \vec{v}\}, \left\{\vec{a}, \vec{b}\right\}$ and $\{\vec{w}, \vec{s}\}$ each constitute an orthogonal basis of $\mathbb{R}^2$:



Notice that all of the vectors start at the origin (black dot). They are separated horizontally for clarity only.

If all of the vectors in an orthogonal basis also have a norm (length) of 1, then the base is called an *orthonormal basis*. A very important example of an orthonormal basis is the *standard basis*.

In $\mathbb{R}^2$ this is the basis composed of the vectors:

$$\hat{e}_1 = \begin{pmatrix}1\\0\end{pmatrix}, \quad \hat{e}_2 = \begin{pmatrix}0\\1\end{pmatrix}.$$

In $\mathbb{R}^3$ the standard basis is composed of the vectors:

$$\hat{e}_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}, \quad \hat{e}_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}, \quad \hat{e}_3 = \begin{pmatrix}0\\0\\1\end{pmatrix}.$$

In $\mathbb{R}^n$ the standard basis is composed of the vectors:

$$\hat{e}_1 = \begin{pmatrix}1\\0\\\vdots\\\vdots\\0\end{pmatrix}, \quad \hat{e}_2 = \begin{pmatrix}0\\1\\\vdots\\\vdots\\0\end{pmatrix}, \quad \dots, \quad \hat{e}_i = \begin{pmatrix}0\\\vdots\\1\\\vdots\\0\end{pmatrix}, \quad \dots, \quad \hat{e}_n = \begin{pmatrix}0\\0\\\vdots\\\vdots\\1\end{pmatrix}.$$

i-th component

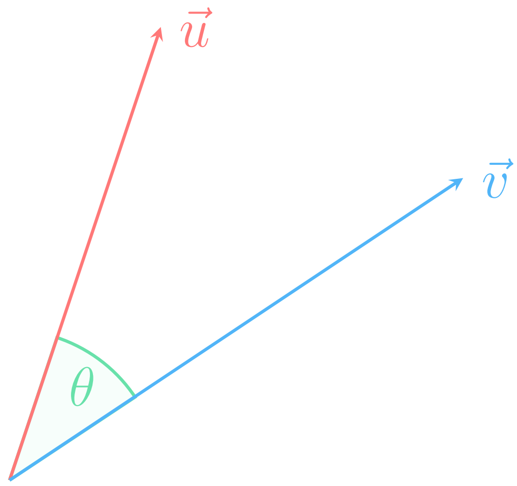
In $\mathbb{R}^2$ and $\mathbb{R}^3$ the basis vectors $\hat{e}_1, \hat{e}_2, \hat{e}_3$ are sometimes called $\hat{\imath}, \hat{\jmath}, \hat{k}$ or $\hat{x}, \hat{y}, \hat{z}$.

[1] *Orthogonal* is a generalization of *perpendicular* for any dimension n - and in fact, any generalized vector space. It is defined more rigourosly in the next subsection.

2.6 Vector-Vector Products

2.6.1 The Dot Product

Since any two **linearly independent** vectors in $\mathbb{R}^n$ span a plane, there is an angle θ between them. This angle can be found using the *dot product* (also called the *scalar product*, and more generaly the *inner product*).



The dot product of $\vec{u}$ and $\vec{v}$ is denoted as $\vec{u} \cdot \vec{v}$ (i.e. with a dot between the vectors, hence the name), and is defined as:

$$\vec{u} \cdot \vec{v} = \|\vec{u}\| \cdot \|\vec{v}\| \cdot \cos(\theta),$$

and thus the angle between two vectors is equal to (by rearranging the previous equation):

$$\theta = \arccos\left(\frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \cdot \|\vec{v}\|}\right).$$

Note 15

Notice that the result of a dot product is a scalar!

Other common notations for the dot product are $\langle \vec{u}, \vec{v} \rangle$ and $\langle \vec{u} \mid \vec{v} \rangle$. The last notation is most common in physics.

Example 54

Let's calculate the angle between the vectors $\vec{u} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ and $\vec{v} = \begin{pmatrix} 5 \\ -1 \end{pmatrix}$. The angle between $\vec{u}$ and the x-axis is

$$\theta_u = \arctan\left(\frac{3}{2}\right) \approx 0.98 \approx 56.3°,$$

while the angle between $\vec{v}$ and the x-axis is

$$\theta_v = \arctan\left(\frac{-1}{5}\right) \approx -0.2 \approx -11.3°.$$

This means that the angle between the two vectors is

$$0.98 + 0.2 = 1.18 \approx 67.6^\circ.$$

The dot product of two vectors is zero, when either one of the vectors is the zero vector (or both of them are), or the angle between the vectors is $\frac{\pi}{2} = 90^\circ$ (since $\cos\left(\frac{\pi}{2}\right) = 0$). This means that when two non-zero vectors have an dot product equaling zero, **the two vectors are orthogonal to each other**.

This is in fact such an important concept, that it deserves a box:

IMPORTANT! 1

When the dot product of two vectors is zero, they are orthogonal (and vice-versa).

When vectors are given with explicit components, the dot product is given by the sum of the element-wise products:

$$\vec{u} \cdot \vec{v} = \sum_{i=1}^{n} u_i \cdot v_i.$$

Example 55

The dot product of the vectors $\vec{u} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$, $\vec{v} = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}$ is:

$$\begin{aligned} \vec{u} \cdot \vec{v} &= 1 \cdot 2 + (0 \cdot -1) + (-1 \cdot 3) \\ &= 2 + 0 - 3 \\ &= -1. \end{aligned}$$

Example 56

Let's calculate again the angle between the vectors $\vec{u} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ and $\vec{v} = \begin{pmatrix} 5 \\ -1 \end{pmatrix}$, this time using the algabreic definition:

$$\begin{aligned} \vec{v} \cdot \vec{u} &= 2 \cdot 5 + 3 \cdot (-1) \\ &= 10 - 3 \\ &= 7. \end{aligned}$$

Furthermore,

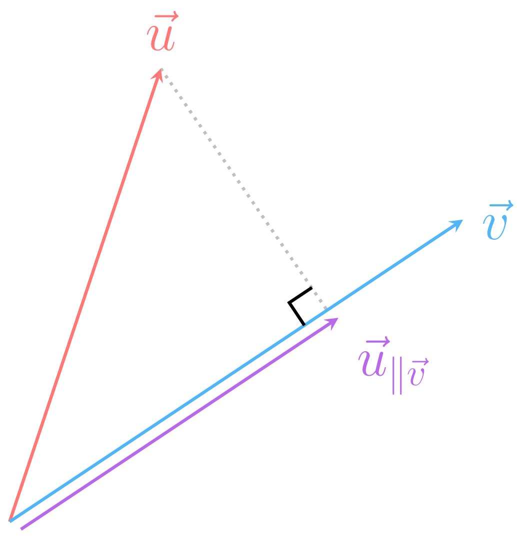
$$\begin{aligned} \|\vec{u}\| &= \sqrt{2^2 + 3^2} = \sqrt{4 + 9} = \sqrt{13} \\ \|\vec{v}\| &= \sqrt{5^2 + (-1)^2} = \sqrt{25 + 1} = \sqrt{26}. \end{aligned}$$

Thus the angle between the vectors is

$$\begin{aligned} \theta &= \arccos\left(\frac{7}{\sqrt{13 \cdot 26}}\right) \\ &\approx 1.18 \approx 67.6^\circ, \end{aligned}$$

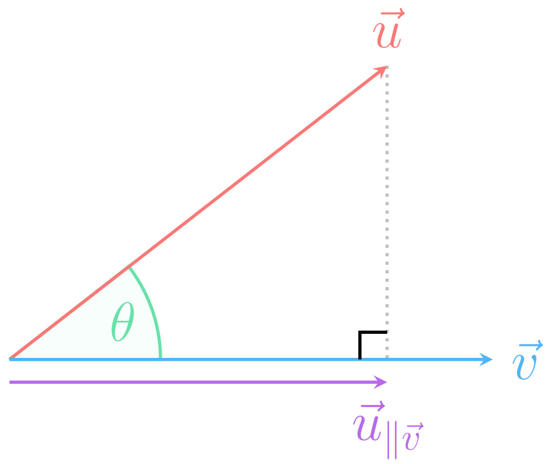
which is the same result we got from the direct calculation.

The dot product between two vectors can be used to calculate the *projection* of a vector on another vector:



In the above figure the vector $\vec{u}$ is projected onto the vector $\vec{v}$, resulting in the purple vector denoted $\vec{u}_{\parallel\vec{v}}$ (which is a bit offset from $\vec{v}$ so both of the vectors can be clearly seen). The projection of $\vec{u}$ on $\vec{v}$ is done by drawing a line from $\vec{u}$ that intersects $\vec{v}$ at a right angle. The resulting projection is then the point of intersection on $\vec{v}$.

The relation of vector projection to the dot product is clearly seen when we rotate the above figure such that $\vec{v}$ lies horizontally:



The length of $\vec{u}_{\parallel\vec{v}}$ is thus

$$\left\|\vec{u}_{\parallel\vec{v}}\right\| = \|\vec{u}\| \cos(\theta) = \frac{\vec{u}\cdot\vec{v}}{\|\vec{v}\|}.$$

Example 57

Write projection calculation example here.

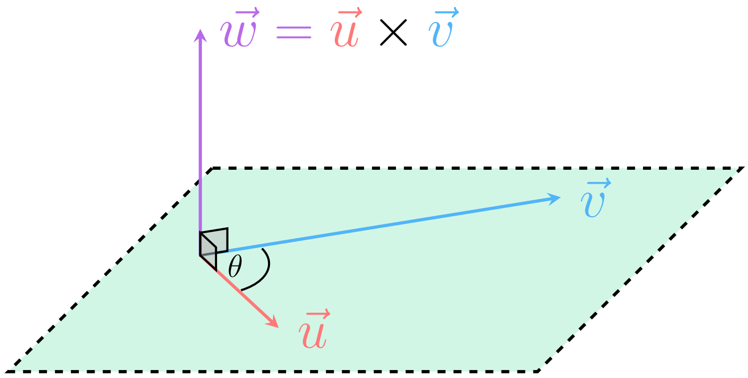
2.6.2 The Cross Product

Another product of two vectors is commonly used: the *cross product*. Unlike the dot product, the cross product is only defined in $\mathbb{R}^3$, and with a somewhat different meaning also in $\mathbb{R}^2$.

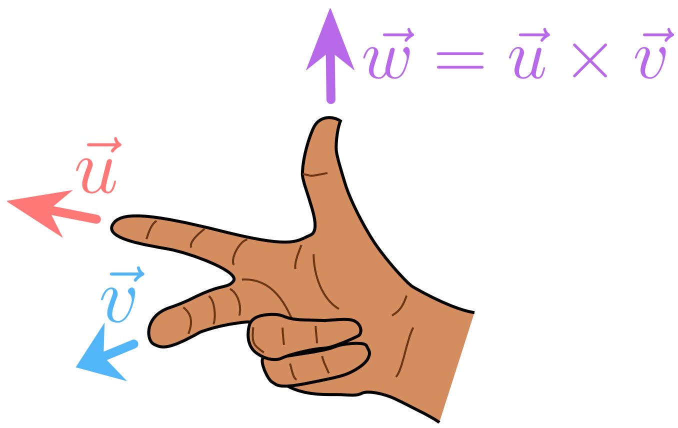
Geometrically, the cross product of two vectors $\vec{u}, \vec{v} \in \mathbb{R}^3$ is defined as a vector $\vec{w}$ which is **orthogonal** to both $\vec{u}$ and $\vec{v}$, and has a magnitude

$$r_w = \|\vec{u}\| \cdot \|\vec{v}\| \cdot \sin(\theta),$$

where θ is the angle between the vectors.



The direction of $\vec{u} \times \vec{v}$ is determined by the *right-hand rule*: using a person's right hand, when $\vec{u}$ points in the direction of the index finger and $\vec{v}$ in the direction of the middle finger, then $\vec{w} = \vec{u} \times \vec{v}$ points in the direction of the thumb:



The cross product is *anti-commutative*, i.e. changing the order of the vectors results in inverting the product:

$$\vec{u} \times \vec{v} = -\vec{v} \times \vec{u}$$

When the vectors are given as column vectors $\vec{u} = \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix}, \vec{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}$, the resulting cross product is

$$\vec{v} \times \vec{w} = \begin{pmatrix} u_y v_z - u_z v_y \\ u_z v_x - u_x v_z \\ u_x v_y - u_y v_x \end{pmatrix}$$

Write ways of memorizing the cross product

Example 58

What is the cross product of $\hat{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and $\hat{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$?

$$\begin{aligned} \hat{e}_1 \times \hat{e}_2 &= \begin{pmatrix} \cancel{0 \cdot 0} - \cancel{0 \cdot 1} \\ \cancel{0 \cdot 0} - \cancel{1 \cdot 0} \\ 1 \cdot 1 - \cancel{0 \cdot 0} \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\ &= \hat{e}_3. \end{aligned}$$

Similarly, $\hat{e}_2 \times \hat{e}_3 = \hat{e}_1$ and $\hat{e}_3 \times \hat{e}_1 = \hat{e}_2$.

Challange 5

Using component calculation and utilizing the dot product, show that $\vec{u} \times \vec{v}$ is indeed orthogonal to both $\vec{u}$ and $\vec{v}$.

3 Linear Transformations

3.1 What is a Linear Transformation?

A *linear transformation* is a function $T : A \to B$, that obeys the follows two criteria:

- For each $x \in A$ and a scalar α:
$$T(\alpha x) = \alpha T(x).$$
- For any $x, y \in A$:
$$T(x + y) = T(x) + T(y).$$

Example 59

The function $f : \mathbb{R} \to \mathbb{R},\ f(x) = 3x$ is linear. Proof by the above criteria:

- For any scalar $\alpha \in \mathbb{R}$,
$$\begin{aligned} f(\alpha x) &= 3(\alpha x) \\ &= 3\alpha x \\ &= \alpha \cdot 3x \\ &= \alpha f(x). \end{aligned}$$
- For any two numbers $x, y \in \mathbb{R}$
$$\begin{aligned} f(x + y) &= 3(x + y) \\ &= 3x + 3y \\ &= f(x) + f(y). \end{aligned}$$

Therefore, f is linear.

Example 60

Is the function $g : \mathbb{R} \to \mathbb{R},\ g(x) = 3x + 5$ linear? Let's check:

- For a scalar $\alpha \in \mathbb{R}$: $g(\alpha x) = 3\alpha x + 5$, but $\alpha g(x) = \alpha(3x + 5) = 3\alpha x + 5\alpha$.
For example, for $x = 1, \alpha = 2$:
$$\begin{aligned} g(\alpha x) &= g(2 \cdot 1) = 6 + 5 = 11 \\ \alpha g(x) &= g(1) = 2(3 + 5) = 2 \cdot 8 = 16 \neq 11. \end{aligned}$$

Since at least one of the criteria is not met (in this case the first criterion), g is **not** linear.

Note 16

For a function to be non-linear, it is **enough** that any one of the two criteria isn't fulfilled. Contrary to that, for a function to be linear, **both** criteria must be fulfilled.

Challange 6

Check if g fulfills the second criterion.

Example 61

Is the function $h : \mathbb{R} \to \mathbb{R},\ h(x) = x^2$ linear? Let's check:

- For $x, y \in \mathbb{R}$: $h(x+y) = (x+y)^2 = x^2 + 2xy + y^2$.
  On the other hand, $h(x) + h(y) = x^2 + y^2$. For $x, y \neq 0$ these are not equal.

Therefore, h is **not** a linear transformation.

Challange 7

Check if h fulfills the first criterion.

The two criteria for linearity can be combined together to a single criterion:

Definition 1

A function $T : A \to B$ is linear, if for any $x, y \in A$ and any scalars α, β:

$$T\left(\alpha x + \beta y\right) = \alpha T(x) + \beta T(y).$$

3.2 Transforming Vectors

Vectors can also be transformed, specifically by functions of the type $T : \mathbb{R}^n \to \mathbb{R}^m$, with $n, m \in \mathbb{N}$. In this course we will mostly concentrate on transformations where the dimensions n, m are each either $1, 2$ or 3 (i.e. transformations of 2- or 3-dimensional vectors, which result in 1-,2- or 3-dimensional vectors). These transformations are much easier to conceptualize (and infintely easier to draw).

Example 62

A transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ is defined as:

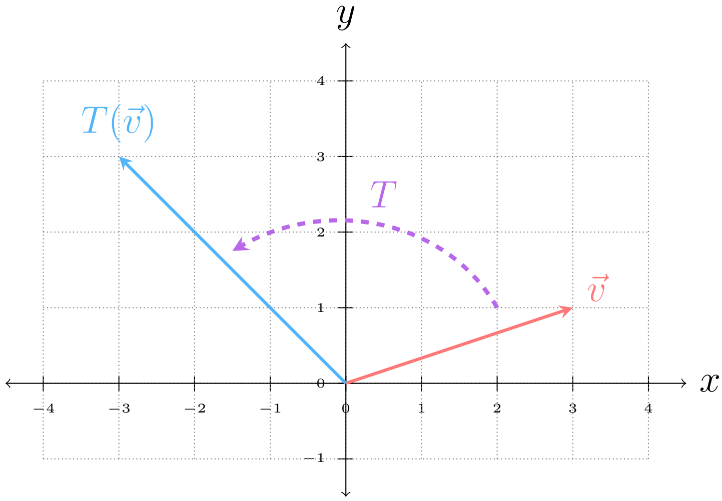
$$T(\vec{v}) = T(v_x, v_y) = \begin{pmatrix} -v_x \\ 3v_y \end{pmatrix},$$

i.e. it takes the vector $\vec{v} = \begin{pmatrix} v_x \\ v_y \end{pmatrix}$, horizontally "flips" its x-component and scales its y-component by 3.

Some examples:

$$\begin{pmatrix}3\\1\end{pmatrix} \xrightarrow{T} \begin{pmatrix}-3\\3\end{pmatrix}, \quad \begin{pmatrix}-1\\2\end{pmatrix} \xrightarrow{T} \begin{pmatrix}1\\6\end{pmatrix}, \quad \begin{pmatrix}-7\\\frac{1}{3}\end{pmatrix} \xrightarrow{T} \begin{pmatrix}7\\1\end{pmatrix}, \quad \begin{pmatrix}0\\0\end{pmatrix} \xrightarrow{T} \begin{pmatrix}0\\0\end{pmatrix}$$

The specific trasformation of the vector $\vec{v} = \begin{pmatrix}3\\1\end{pmatrix}$ looks graphically as follows:



Note 17

Remember that vectors are always drawn starting from the origin!

3.3 Transforming Spaces

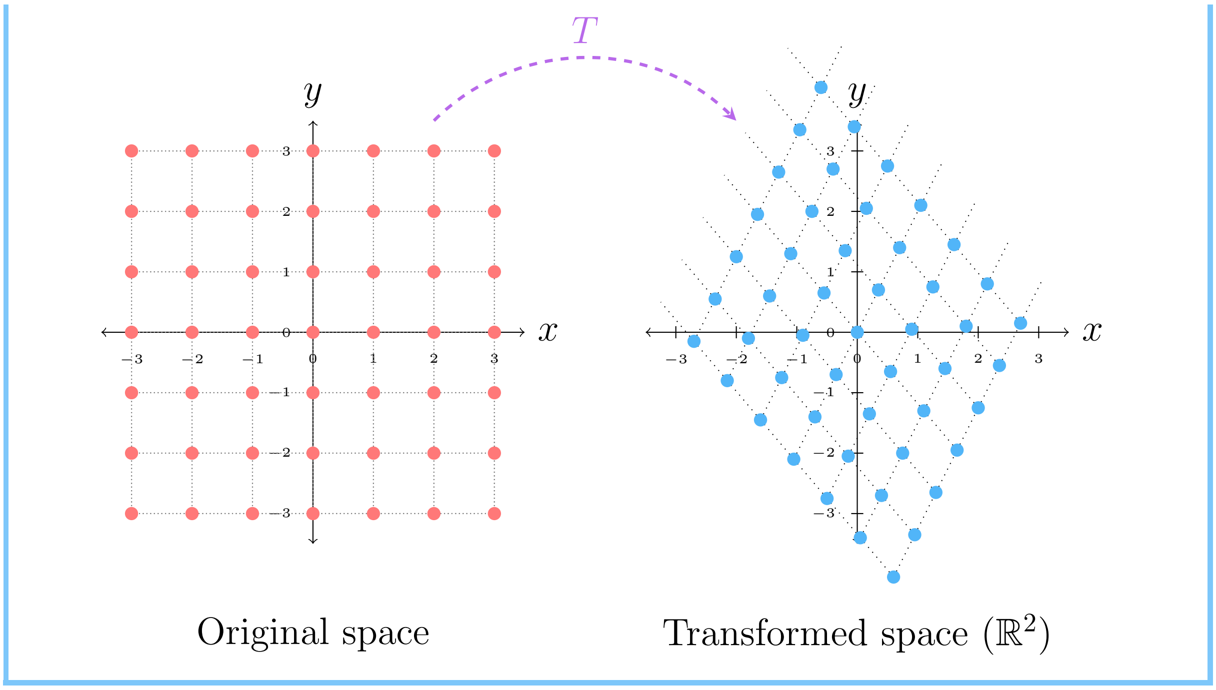
To visualize the way the entire space transforms by a transformation T, we can look at many points transforming together. Since each point represents a vector from the origin to that point, this will show us how different vectors are transformed. For this purpose, we can take the intersections of all the grid lines, then transforming them (here shown in blue), and draw the grid lines between them.

Example 63

The transformation T is defined by its action on a vector $\begin{pmatrix}x\\y\end{pmatrix}$ as follows:

$$T\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}0.35x - 0.55y\\0.7x + 0.65y\end{pmatrix}.$$

Graphically, this is how the transformation affects $\mathbb{R}^2$ (the red dots on the left are lying on the integer grid points, and are transformed to the blue dots on the right):



The example above is a linear transformation. Quick proof:

- T is scalable:

$$T\left(\alpha\begin{pmatrix}x\\y\end{pmatrix}\right)=T\begin{pmatrix}\alpha x\\\alpha y\end{pmatrix}=\begin{pmatrix}0.35\alpha x-0.55\alpha y\\0.7\alpha x+0.65\alpha y\end{pmatrix}=\begin{pmatrix}\alpha\,(0.35x-0.55y)\\\alpha\,(0.7x+0.65y)\end{pmatrix}$$
$$=\alpha\begin{pmatrix}0.35x-0.55y\\0.7x+0.65y\end{pmatrix}=\alpha T\begin{pmatrix}x\\y\end{pmatrix}.$$

- T is separable:

$$T\left(\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}a\\b\end{pmatrix}\right)=T\begin{pmatrix}x+a\\y+b\end{pmatrix}=\begin{pmatrix}0.35\,(x+a)-0.55\,(y+b)\\0.7\,(x+a)+0.65\,(y+b)\end{pmatrix}=\begin{pmatrix}0.35x+0.35a-0.55y-0.55b\\0.7x+0.7a+0.65y+0.65b\end{pmatrix}$$
$$=\begin{pmatrix}0.35x-0.55y\\0.7x+0.65y\end{pmatrix}+\begin{pmatrix}0.35a-0.55b\\0.7a+0.65b\end{pmatrix}=T\begin{pmatrix}x\\y\end{pmatrix}+T\begin{pmatrix}a\\b\end{pmatrix}.$$

3.4 Characteristics of Linear Transformations

The graphical representation of the effect of T on $\mathbb{R}^2$ above reveals several important characteristics of linear transformations:

- Linear transformations always preserve the zero vector, i.e. they always map $\vec{0}$ to $\vec{0}$ (in the case of $\mathbb{R}^2$: $\begin{pmatrix}0\\0\end{pmatrix}\xrightarrow{T}\begin{pmatrix}0\\0\end{pmatrix}$).
- Parallel lines remain parallel after the transformation is applied.
- While areas can scale under the transformation, the ratio between areas is preserved.

Challange 8

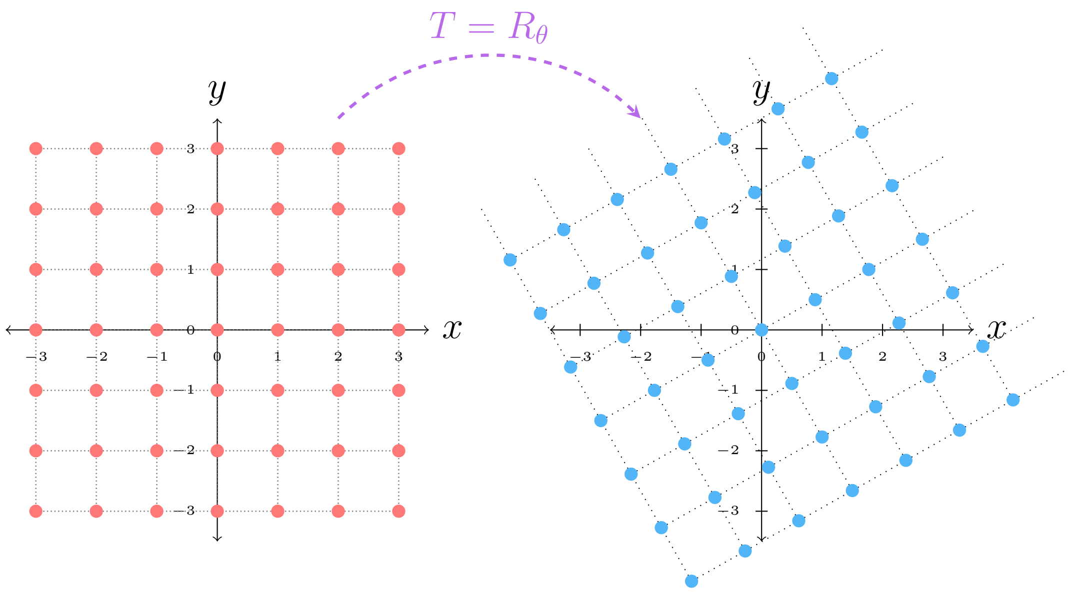
Show that all three characteristics above can be derived directly from the definitions of a linear transformation.

3.5 Basic Types of Linear Trasformations

In $\mathbb{R}^2$ there are several basic linear transformation from which all other linear transformations are constructed (via composition).

3.5.1 Rotations Around the Origin

Rotating space around its origin counter clockwise by an angle θ. The transformation is usually denoted as R_θ.

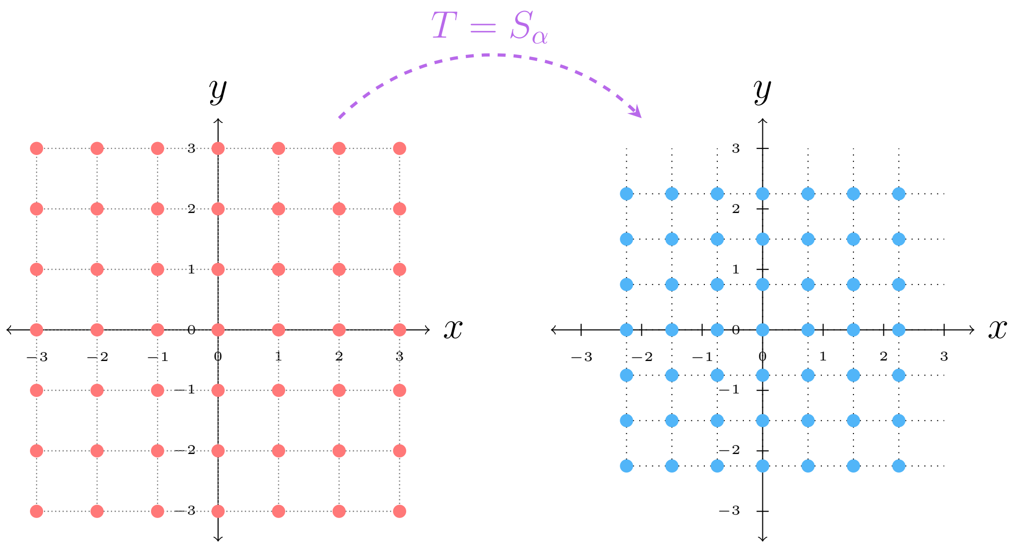


3.5.2 Scaling

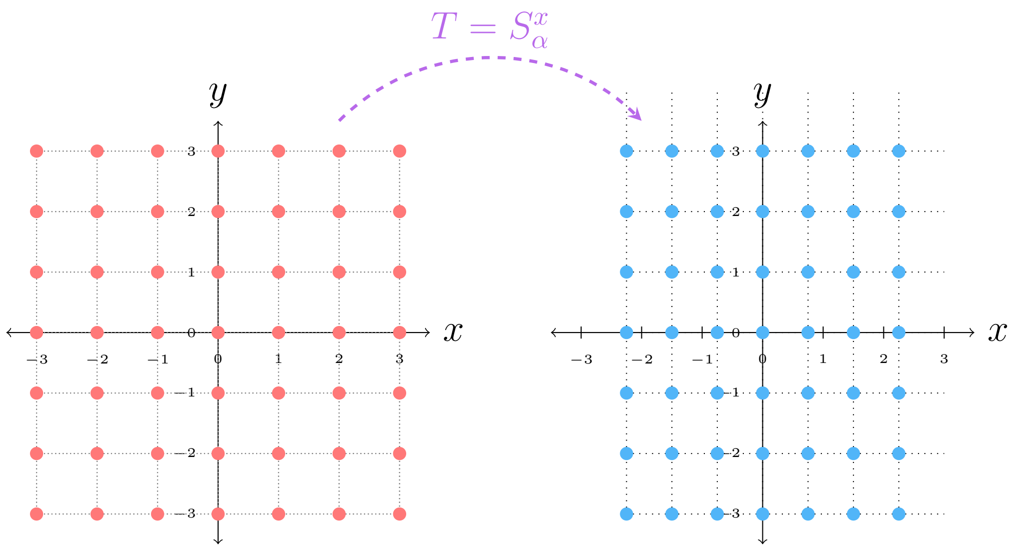
The scaling of space by a scalar is a linear translation. In addition, scaling space by a scalar in either the x or y direction is also a linear transformation. We can denote a scaling transformation by S_α, where α is the scalar. Similarily, scaling in the x direction would be S^x_α, and in the y direction S^y_α.

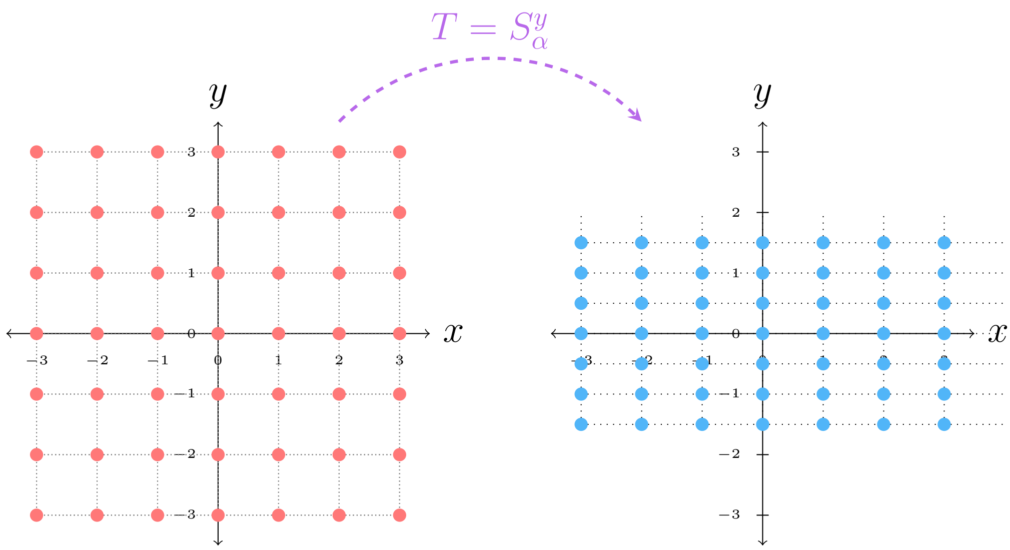
Note 18

In the following graphical representations all scalings are done by scalars smaller than 1 for graphical reasons, but scalings can of course be done for any scalar $\alpha \in \mathbb{R}$).



Scaling in one direction only:





TODO: ADD MORE TRANSFORMATIONS

Note 19

Addition of some constant non-zero vector, known as a *translation*, is **not** a linear transformation. These transformations are a type of *affine* transformations which are not discussed in this course, and can be represented as a linear transformation in higher dimensions (as we will see later in the course).

4 Matrices

4.1 Matrices from Linear Transformations

A vector can be written as the linear combination of its components. These components can then be represented as a scalar α_i multiplied by the respective basis vector $\hat{e}_i$.

Example 64

The vector $\vec{v} = \begin{pmatrix} -1 \\ 2 \\ 5 \end{pmatrix}$ can be written as

$$\begin{aligned}
\vec{v} &= \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 5 \end{pmatrix} \\
&= -1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 5 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\
&= -1\hat{e}_1 + 2\hat{e}_2 + 5\hat{e}_3 \\
&= -1\hat{x} + 2\hat{y} + 5\hat{z}.
\end{aligned}$$

The last form is pretty standard when using $\mathbb{R}^3$, especially in phyiscs.

When a linear transformation T is applied to a vector $\vec{v} = \alpha_1\hat{e}_1 + \alpha_2\hat{e}_2 + \cdots + \alpha_n\hat{e}_n$, linearity dictates the follows:

$$\begin{aligned}
T(\vec{v}) &= T\left(\alpha_1\hat{e}_1 + \alpha_2\hat{e}_2 + \cdots + \alpha_n\hat{e}_n\right) \\
&= T\left(\alpha_1\hat{e}_1\right) + T\left(\alpha_2\hat{e}_2\right) + \cdots + T\left(\alpha_n\hat{e}_n\right) \\
&= \alpha_1 T\left(\hat{e}_1\right) + \alpha_2 T\left(\hat{e}_2\right) + \cdots + \alpha_n T\left(\hat{e}_n\right).
\end{aligned}$$

This means that the transformed vector is a linear combination of the transformed basis using the same components as before the transformation.

Example 65

$\vec{v} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$, $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -2x + y \\ 3x - 2y \end{pmatrix}$. Normally we would write

$$T(\vec{v}) = T\begin{pmatrix} 3 \\ -1 \end{pmatrix} = \begin{pmatrix} -6 - 1 \\ 9 + 2 \end{pmatrix} = \begin{pmatrix} -7 \\ 11 \end{pmatrix}.$$

Instead, let's see how the basis vectors $\hat{e}_1, \hat{e}_2$ transform:

$$T(\hat{e}_1) = \begin{pmatrix} -2+0 \\ 3+0 \end{pmatrix} = \begin{pmatrix} -2 \\ 3 \end{pmatrix}$$
$$T(\hat{e}_2) = \begin{pmatrix} 0+1 \\ 0-2 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}.$$

Then, we write $T(\vec{v})$ as a linear combination of these transformed basis vectors:

$$\begin{aligned} T(\vec{v}) = 3T(\hat{e}_1) - T(\hat{e}_2) &= 3\begin{pmatrix} -2 \\ 3 \end{pmatrix} - \begin{pmatrix} 1 \\ -2 \end{pmatrix} \\ &= \begin{pmatrix} -6 \\ 9 \end{pmatrix} - \begin{pmatrix} 1 \\ -2 \end{pmatrix} = \begin{pmatrix} -6-1 \\ 9+2 \end{pmatrix} = \begin{pmatrix} -7 \\ 11 \end{pmatrix}. \end{aligned}$$

This is, of course, the exact same result we got from the direct calculation.

If we look at a generic transformation T (for the sake of simplicity $T : \mathbb{R}^2 \to \mathbb{R}^2$) acting on a generic vector $\vec{v} = \begin{pmatrix} x \\ y \end{pmatrix}$,

$$T(\vec{v}) = T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax+by \\ cx+dy \end{pmatrix},$$

we can see a nice pattern: the first component of the transformed vector is $ax + by$, which is a dot product of some vector $\begin{pmatrix} a \\ b \end{pmatrix}$ with $\begin{pmatrix} x \\ y \end{pmatrix}$. Similarly, the second component of the transformed vector is $cx + dy$, which is a dot product of some vector $\begin{pmatrix} c \\ d \end{pmatrix}$ with $\begin{pmatrix} x \\ y \end{pmatrix}$.

We can write this observation in a compact way as a *matrix* (plural: *matrices*) we will call M:

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

and we define the product of a matrix M with a vector $\begin{pmatrix} x \\ y \end{pmatrix}$ as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax+by \\ cx+dy \end{pmatrix}.$$

In this sense, matrices are a compact way of writing linear transformations performed on vectors.

4.2 Basics of Matrices

In a more general sense, a matrix is a collection of numbers (refered to as *elements*, and are usually real or complex) that are sorted into *rows* and *columns*.

Example 66

The follows matrix A has 3 rows and 2 columns, and is therefore a 3×2 matrix:

$$A = \begin{pmatrix} 3 & 7 \\ 1 & -2 \\ 0 & -5 \end{pmatrix}$$

We denote the elements of a matrix as undercase latin letters, with 2 indices that represent the element's row and column.

Example 67

The above matrix has the follows elements:

$$\begin{aligned} a_{11} &= 3, & a_{12} &= 7, \\ a_{21} &= 1, & a_{22} &= -2, \\ a_{31} &= 0, & a_{32} &= -5. \end{aligned}$$

Note 20

The notation for matrix elements might cause confusion, for example reading b_{12} as "bee twelve" (when in fact it is "bee one-two", i.e. first row, second column of the matrix B). Always pay attention to the context and what makes sense within it.

A general $M \times N$ matrix looks as follows (indices in red represent the row number of an element, while indices in blue represent its column number):

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \dots \\ a_{M1} & a_{M2} & \dots & a_{MN} \end{pmatrix}$$

A matrix for which $M = N$ (i.e. it has the same number of rows and columns) is called a *square matrix*. In a square matrix, the elements where the two indices are equal (i.e. $i = j$) are said to be found on the *main diagonal* of the matrix (also called the *major diagonal*, the *principal diagonal*, and the *primary diagonal*).

Example 68

In the follows matrices, the main diagonal elements are highlighted:

$$A = \begin{pmatrix} 4 & -7 & 1 & 0 \\ 1 & 5 & -1 & 3 \\ 3 & -7 & 1 & 6 \\ -5 & 9 & 12 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & -1 \\ 1 & 1 & 0 \end{pmatrix}$$

Also in a square matrix ($N \times N$ matrix), the elements a_{ij} for which $j = N - i + 1$ are said to be found on the *antidiagonal* (also called the *minor diagonal*, the *counter diagonal*, and the *secondary diagonal*)

Example 69

In the follows matrices, the anti-main diagonal elements are highlighted:

$$A = \begin{pmatrix} 3 & -1 & 5 & 7 \\ 0 & 2 & -4 & 2 \\ 4 & 0 & -3 & 2 \\ 1 & 3 & 4 & 5 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 0 & -3 \\ 5 & 1 & 3 \\ 9 & 7 & -6 \end{pmatrix}$$

4.3 Diagonal and Triangular Matrices

A matrix with all elements outside the main diagonal equaling zero is called a *diagonal matrix*.

Example 70

The follows matrices are diagonal matrices:

$$A = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -6 \end{pmatrix}$$

Note 21

Notice that the main diagonal elements in a diagonal matrix **could be** equal to zero. The only creterion is that the elements that are not on the diagonal **must be** zero.

A matrix with all elements "above" the diagonal equaling zero is called an *upper triangular matrix*. Similarily, a matrix with all elements "below" the main diagonal equaling zero is called a *lower triangular matrix*.

Example 71

An upper triangular matrix:

$$A = \begin{pmatrix} 5 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ -3 & 4 & 7 & 0 \\ 2 & 2 & 0 & -1 \end{pmatrix}$$

A lower triangular matrix:

$$B = \begin{pmatrix} 5 & 7 & 9 & 1 \\ 0 & 3 & 4 & -3 \\ 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

4.4 Transpose of a Matrix

Transposing a matrix is a common operation. When transposing a matrix A (notated by $A^\top$), we take the elements and exchange their row with their column:

$$a_{ij} \xrightarrow{\text{transpose}} a_{ji}$$

Example 72

$$\begin{pmatrix} 2 & 1 & 3 \\ 0 & -7 & 1 \\ 4 & 5 & 0 \end{pmatrix} \xrightarrow{\top} \begin{pmatrix} 2 & 0 & 4 \\ 1 & -7 & 5 \\ 3 & 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 5 \\ 4 & 3 & 2 \end{pmatrix} \xrightarrow{\top} \begin{pmatrix} 1 & 4 \\ 0 & 3 \\ 5 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 2 & -1 & 3 & 7 \end{pmatrix} \xrightarrow{\top} \begin{pmatrix} 2 \\ -1 \\ 3 \\ 7 \end{pmatrix}$$

The elements with identical row and column numbers (i.e. $a_{11}, a_{22}, a_{33}, \dots$) remain in their previous place. In the case of a square matrix, these are the elements on the main diagonal.

Example 73

$$\begin{pmatrix} \boxed{2} & 1 & 3 \\ 0 & \boxed{-7} & 1 \\ 4 & 5 & \boxed{0} \end{pmatrix} \xrightarrow{\top} \begin{pmatrix} \boxed{2} & 0 & 4 \\ 1 & \boxed{-7} & 5 \\ 3 & 1 & \boxed{0} \end{pmatrix}$$

Of course, the transpose of a transposed matrix is the original matrix:

$$\left(A^{\top}\right)^{\top} = A,$$

and generally the transpose of an $M \times N$ matrix is an $N \times M$ matrix. Specifically, a transpose of a row vector is a column vector and vice-versa.

4.5 Adding Two Matrices

Addition of matrices is only allowed for matrices of the same exact size, and is done **element wise** (exactly like addition of vectors): for the matrices A, B, C with elements $\{a_{ij}\}, \{b_{ij}\}, \{c_{ij}\}$ respectively,

$$c_{ij} = a_{ij} + b_{ij}.$$

Example 74

$$\begin{pmatrix} 1 & 5 & 10 \\ -3 & 4 & 0 \end{pmatrix} + \begin{pmatrix} 2 & 0 & -7 \\ 3 & 9 & -1 \end{pmatrix} = \begin{pmatrix} 3 & 5 & 3 \\ 0 & 13 & -1 \end{pmatrix}$$

$$\begin{pmatrix} 4 & 1 & 3 & 0.5 \\ -3 & 13 & -2 & -1 \\ 2 & 1 & -1 & 5 \\ 7 & -5 & -1 & 0 \end{pmatrix} + \begin{pmatrix} -1 & 1 & 2 & 2 \\ 0 & 3 & -5 & 3 \\ 3 & 3 & 4 & 5 \\ 4 & 7 & 9 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 2 & 5 & 2.5 \\ -3 & 16 & -7 & -2 \\ 5 & 4 & 3 & 10 \\ 11 & 2 & 8 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 2 \\ 5 \\ 9 \\ 1 \\ 4 \end{pmatrix} + \begin{pmatrix} 0 \\ -0.5 \\ 9 \\ 2 \\ -3 \end{pmatrix} = \begin{pmatrix} 2 \\ 4.5 \\ 18 \\ 3 \\ 1 \end{pmatrix}$$

4.6 Multiplying a Matrix by a Scalar

Similarily to vectors, matrices can be multiplied by a scalar. This is done by multiplying each element of the matrix by that same scalar.

Example 75

$$3 \cdot \begin{pmatrix} 1 & 2 & -4 \\ 5 & 7 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 6 & -12 \\ 15 & 21 & 3 \end{pmatrix}$$

$$-2 \cdot \begin{pmatrix} 0 & 5 & 1 \\ 2 & -3 & 4 \\ 6 & 3 & 7 \end{pmatrix} = \begin{pmatrix} 0 & -10 & -2 \\ -4 & 6 & -8 \\ -12 & -6 & -14 \end{pmatrix}$$

4.7 Multiplying a Matrix and a Vector

As discussed briefly above, multiplying a matrix and a vector results in applying some linear transformation to the vector. This is of course possible only when the number of **columns** in the matrix is the same as the number of elements of the vector.

Example 76

The follows matrix-vector multiplication are **allowed**:

$$\begin{pmatrix} 5 & 3 & 1 \\ 4 & -1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ -1 \\ 4 \end{pmatrix} \qquad \begin{pmatrix} 1 & 5 \\ 5 & 0 \\ -3 & 2 \\ 4 & 7 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -5 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 7 \\ 5 \\ 1 \end{pmatrix}$$

The follows matrix-vector multiplication are **NOT** allowed:

$$\begin{pmatrix} 5 & 3 & 1 \\ 4 & -1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 5 \\ 5 & 0 \\ -3 & 2 \\ 4 & 7 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -5 \\ 0 \\ 2 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 7 \\ 1 \end{pmatrix}$$

We can interpret the **columns** of the matrix as showing as how the standard basis vectors are transformed.

Notice that in all of the examples above, the matrix is always **to the left** of the vector. This is a standard way to do matrix-vector multiplication when the vectors are **column vectors**. When they are **row vectors**, we usually position the matrix **to the right** of the vector. This will become more obvious when we discuss matrix-matrix multiplication soon.

When multiplying an $M \times N$ matrix by an N-dimensional vector, the result is a vector of M-dimensions, with its i-th element being equal to the dot product of the i-th row of the matrix and the original vector.

Example 77

$$\begin{pmatrix} 3 & -1 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} 5 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \cdot 5 + (-1) \cdot 0 \\ 1 \cdot 5 + 5 \cdot 0 \end{pmatrix} = \begin{pmatrix} 15 \\ 5 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 4 & 0 \\ 6 & -3 & 7 \\ 4 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 6 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 + 4 \cdot 6 + 0 \cdot 3 \\ 6 \cdot 1 + (-3) \cdot 6 + 7 \cdot 3 \\ 4 \cdot 1 + 2 \cdot 6 + 1 \cdot 3 \end{pmatrix} = \begin{pmatrix} 1 + 24 + 0 \\ 6 - 18 + 21 \\ 4 + 12 + 3 \end{pmatrix} = \begin{pmatrix} 25 \\ 9 \\ 18 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 1 \\ 3 & -4 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 5 \\ -2 \end{pmatrix} = \begin{pmatrix} 1 \cdot 0 + 2 \cdot 5 + 1 \cdot (-2) \\ 3 \cdot 0 + (-4) \cdot 5 + 1 \cdot (-2) \end{pmatrix} = \begin{pmatrix} 0 + 10 - 2 \\ 0 - 20 - 2 \end{pmatrix} = \begin{pmatrix} 8 \\ -22 \end{pmatrix}$$

$$\begin{pmatrix} 5 & 2 \\ 1 & -1 \\ 3 & 7 \\ 6 & 0 \end{pmatrix} \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \cdot 3 + 2 \cdot 1 \\ 1 \cdot 3 + (-1) \cdot 1 \\ 3 \cdot 3 + 7 \cdot 3 \\ 6 \cdot 3 + 0 \cdot 1 \end{pmatrix} = \begin{pmatrix} 15 + 2 \\ 3 - 1 \\ 9 + 21 \\ 18 + 0 \end{pmatrix} = \begin{pmatrix} 17 \\ 2 \\ 30 \\ 18 \end{pmatrix}$$

The last two multiplications above show us that a linear transformation of a vector via matrix multiplication can result in a new vector of different dimensionality. Generally, the transformation represented by a multiplication of a real $M \times N$ matrix and a real N-dimensional vector is of the type $T : \mathbb{R}^N \to \mathbb{R}^M$.

Let us show that a multiplication of any 2×2 matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and a 2-dimensional vector $\vec{v} = \begin{pmatrix} x \\ y \end{pmatrix}$ is indeed a linear transformation:

- $T(\beta x) = \beta T(x)$ (we use β for a generic scalar as to avoid confusion with the matrix elements $\{a_{ij}\}$):

$$\begin{aligned} A \cdot \beta \vec{v} &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \beta x \\ \beta y \end{pmatrix} \\ &= \begin{pmatrix} \beta a_{11} x + \beta a_{12} y \\ \beta a_{21} x + \beta a_{22} y \end{pmatrix} \\ &= \begin{pmatrix} \beta \left(a_{11} x + a_{12} y \right) \\ \beta \left(a_{21} x + a_{22} y \right) \end{pmatrix} \\ &= \beta \begin{pmatrix} a_{11} x + a_{12} y \\ a_{21} x + a_{12} y \end{pmatrix} \\ &= \beta \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \beta A \vec{v}. \end{aligned}$$

- $T(x + y) = T(x) + T(y)$:

$$\begin{aligned} A \cdot (\vec{v}_1 + \vec{v}_2) &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right) \\ &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix} \\ &= \begin{pmatrix} a_{11} \left(x_1 + x_2 \right) + a_{12} \left(y_1 + y_2 \right) \\ a_{21} \left(x_1 + x_2 \right) + a_{22} \left(y_1 + y_2 \right) \end{pmatrix} \\ &= \begin{pmatrix} a_{11} x_1 + a_{11} x_2 + a_{12} y_1 + a_{12} y_2 \\ a_{21} x_1 + a_{21} x_2 + a_{22} y_1 + a_{22} y_2 \end{pmatrix} \\ &= \begin{pmatrix} a_{11} x_1 + a_{12} y_1 \\ a_{21} x_1 + a_{22} y_1 \end{pmatrix} + \begin{pmatrix} a_{11} x_2 + a_{12} y_2 \\ a_{21} x_2 + a_{22} y_2 \end{pmatrix} \end{aligned}$$

$$= A\vec{v}_1 + A\vec{v}_2.$$

Challange 9

Show that any $M \times N$ matrix multiplied by an N-dimensional vector is indeed a linear transformation.

4.8 Important Matrices

4.8.1 The Identity Matrix

Each dimension $n = 1, 2, 3, \dots$ has an identity matrix, written as I_n. The identity matrix has ones on its main diagonal, and zeros anywhere else:

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},\ \cdots,\ I_n = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

The identity matrix when multiplied with a vector does not change the vector at all (hence its name).

Example 78

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -3 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \cdot (-3) + \cancel{0 \cdot (2)} \\ \cancel{0 \cdot (-3)} + 1 \cdot 2 \end{pmatrix} = \begin{pmatrix} -3 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 5 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 1 \cdot 5 + \cancel{0 \cdot 1} + \cancel{0 \cdot (-2)} \\ \cancel{0 \cdot 5} + 1 \cdot 1 + \cancel{0 \cdot (-2)} \\ \cancel{0 \cdot 5} + \cancel{0 \cdot 1} + 1 \cdot (-2) \end{pmatrix} = \begin{pmatrix} 5 \\ 1 \\ -2 \end{pmatrix}$$

A short-hand way of writing the identity matrix is by using the *Kronecker delta*. It is defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Thus, we can say that the identity matrix is simply $I = \{\delta_{ij}\}$ (where as always, i is the row index and j is the column index).

4.8.2 Rotation Matrices

In $\mathbb{R}^2$ the rotation matrix $\operatorname{Rot}(\theta)$ rotates the space by θ **counter clock-wise** around the origin. It is defined as:

$$\operatorname{Rot}(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

Example 79

Rotation by 90° is done by

$$\mathrm{Rot}(90°) = \begin{pmatrix} \cos(90°) & -\sin(90°) \\ \sin(90°) & \cos(90°) \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Acting on a the vector $\vec{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ it yeilds

$$\mathrm{Rot}(90°) \cdot \vec{v} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \cancel{0 \cdot 1} - 1 \cdot 1 \\ 1 \cdot 1 + \cancel{0 \cdot 1} \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

which is indeed orthogonal to $\vec{v}$:

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix} = 1 \cdot (-1) + 1 \cdot 1 = -1 + 1 = 0.$$

Challange 10

Show that applying $\mathrm{Rot}(90°)$ to a vector $\vec{v}$ indeed yields a vector that is orthogonal to $\vec{v}$, i.e.

$$\big(\mathrm{Rot}(90°) \cdot \vec{v}\big) \cdot \vec{v} = 0.$$

In $\mathbb{R}^3$ there are 3 standard rotation matrix: one for rotation around each axis (x, y, z):

$$R^x_\theta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix}, \quad R^y_\theta = \begin{pmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix}, \quad R^z_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Notice that for each of these matrix, the corresponding row and column $(x = 1, y = 2, z = 3)$ are each the respective standard basis, and the rest of the matrix is simply $\mathrm{Rot}(\theta)$ in $\mathbb{R}^2$ (with the small caviat that for R^y_θ the rotation is in the opposite direction).

4.8.3 Scaling Matrices

A matrix of the type

$$S_\alpha = \begin{pmatrix} \alpha & 0 & \dots & 0 \\ 0 & \alpha & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha \end{pmatrix},$$

i.e. a diagonal matrix with the diagonal elements are all equal to α, performs a scaling of the space by the parameter α.

Example 80

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

This action should be somewhat obvious, as $S_\alpha = \alpha I_n$.

Scaling can generally be done for each axis separately by multiplying each column of the identity matrix by a different scaling factor:

$$S = \begin{pmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha_n \end{pmatrix},$$

which scales vectors in the x-axis by α_1, in the y-axis by α_2, etc. As can be seen, essentially any diagonal matrix is a scaling matrix.

Example 81

The matrix $S = \begin{pmatrix} 3 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$ scales vectors by 3 in the x-axis and and $\frac{1}{2}$ in the y-axis. For example,

$$\begin{pmatrix} 3 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}.$$

4.8.4 Reflection Matrices

A reflection is a linear transformation only if it's done across a line going through the origin.

Reflecting across the x-axis ($0°$) is represented by the matrix

$$\mathrm{Ref}(0°) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

as it flips $\hat{y}$: transforming it from $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ to $\begin{pmatrix} 0 \\ -1 \end{pmatrix}$.

Similarily, reflecting across the y-axis ($90°$) is represented by the matrix

$$\mathrm{Ref}(90°) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},$$

which flips $\hat{x}$.

Reflecting across the line $y = x$ (which angled by $\theta = 45°$ relative to the x-axis) is represented by the matrix

$$\mathrm{Ref}(45°) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

which can be thought of as exchanging the base vectors $\hat{x}$ and $\hat{y}$.

A general reflection across a line $y = mx$, where $m = \arctan(\theta)$ is represented by

$$\mathrm{Ref}(\theta) = \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix}.$$

Challange 11

Show that reflection across a line that doesn't go through the origin is **not** a linear transformation.

4.8.5 Skew Matrices

A *skew matrix* (also known as a *shear matrix* is a matrix that skews space by some amount k in one direction (e.g. in the case of $\mathbb{R}^2$ either $\hat{x}$ or $\hat{y}$).

In $\mathbb{R}^2$, A skew matrix in the x-direction has the form

$$\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix},$$

where k is the skew parameter.

Similarly, a skew matrix in the y-direction is given by

$$\begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix},$$

where, again, k is the skew parameter.

4.9 Multiplying Two Matrices

We can consider a matrix as a collection of column vectors.

Example 82

The following matrix can be considered a collection of 3 colum vectors, each of 4-dimensions:

$$\begin{pmatrix} \boxed{\begin{matrix}1\\3\\2\\0\end{matrix}} & \boxed{\begin{matrix}5\\4\\1\\-2\end{matrix}} & \boxed{\begin{matrix}7\\0\\-5\\3\end{matrix}} \end{pmatrix}$$

With this in mind, matrix-matrix multiplication $A \cdot B$ can be viewed as follows: we multiply A by each column vector in B separately, and collecting the resulting vectors as columns in a new matrix C.

Example 83

Consider the follows matrix-matrix product:

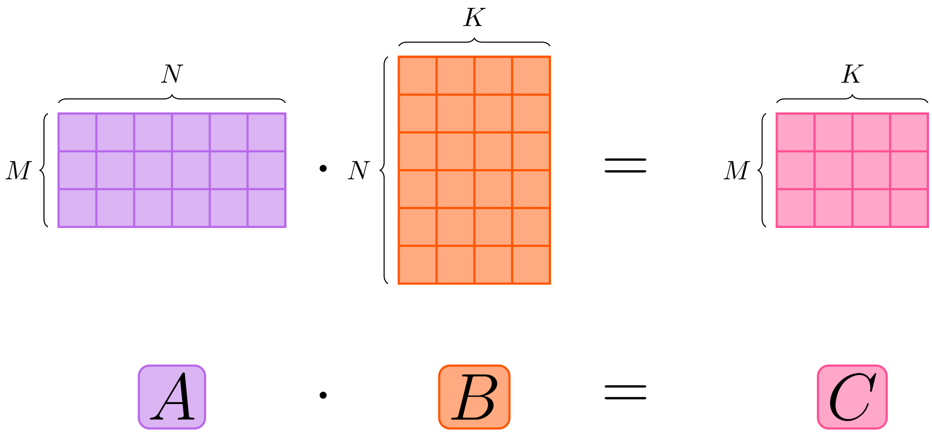
$$A \cdot B = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 4 & 5 \end{pmatrix}$$

We now look at the product of the matrix A with each column of B separately (we will call these columns $\vec{B}_1$ and $\vec{B}_2$):

$$A \cdot \vec{B}_1 = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \cdot 0 + 2 \cdot 4 \\ 0 \cdot 0 + 3 \cdot 4 \end{pmatrix} = \begin{pmatrix} 8 \\ 12 \end{pmatrix}$$
$$A \cdot \vec{B}_2 = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 + 2 \cdot 5 \\ 0 \cdot 4 + 3 \cdot 5 \end{pmatrix} = \begin{pmatrix} 11 \\ 15 \end{pmatrix}$$

Thus, the result of $A \cdot B$ is $C = \begin{pmatrix} 8 & 11 \\ 12 & 15 \end{pmatrix}$.

We can see that there is a restriction on matrix-matrix multiplication: the number of *columns* in the left matrix must be equal to the number of *rows* in the right matrix. The multiplication results in a new matrix with the same number of rows as the left matrix, and same number of columns as the right matrix.



Example 84

Allowed matrix-matrix multiplications:

$$\begin{pmatrix} 1 & 3 & -2 \\ -3 & 5 & 0 \end{pmatrix} \begin{pmatrix} 1 & 3 & -2 & 0 \\ -3 & 5 & 0 & 7 \\ -4 & 2 & 11 & 3 \end{pmatrix} \qquad \begin{pmatrix} 5 & -1 \\ 7 & 3 \end{pmatrix} \begin{pmatrix} 7 & 4 & 0 \\ 5 & 3 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 0 & 1 & 7 \end{pmatrix} \begin{pmatrix} 0 & 1 & -1 \\ 2 & 2 & 0 \\ 0 & 3 & -5 \end{pmatrix}$$

Not allowed matrix-matrix multiplications:

$$\begin{pmatrix} 1 & 3 & -2 \\ -3 & 5 & 0 \end{pmatrix} \begin{pmatrix} 1 & 3 & -2 \\ -4 & 2 & 11 \end{pmatrix} \qquad \begin{pmatrix} 5 & -1 \\ 7 & 3 \end{pmatrix} \begin{pmatrix} 7 & 0 \\ 2 & -1 \\ 4 & 0 \\ 5 & \frac{1}{2} \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 0 & 1 & 7 \end{pmatrix} \begin{pmatrix} 0 & 1 & -1 & 13 \\ 2 & 2 & 0 & -7 \\ 5 & 3 & 1 & 1 \\ 0 & 3 & -5 & 9 \end{pmatrix}$$

Example 85

$$\begin{pmatrix} 1 & 0 & 1 & 2 \\ 2 & 3 & 5 & 0 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 3 \\ 0 & 5 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 + 0 \cdot 2 + 1 \cdot 0 + 2 \cdot (-1) & 1 \cdot 3 + 0 \cdot 3 + 1 \cdot 5 + 2 \cdot 2 \\ 2 \cdot 1 + 3 \cdot 2 + 5 \cdot 0 + 0 \cdot (-1) & 2 \cdot 3 + 3 \cdot 3 + 5 \cdot 5 + 0 \cdot 2 \end{pmatrix} = \begin{pmatrix} -1 & 12 \\ 8 & 40 \end{pmatrix}$$

$$\begin{pmatrix} -3 & 9 \\ 5 & 4 \end{pmatrix} \begin{pmatrix} 1 & 6 \\ 4 & 0 \end{pmatrix} = \begin{pmatrix} (-3) \cdot 1 + 9 \cdot 4 & (-3) \cdot 6 + 9 \cdot (0) \\ 5 \cdot 1 + 4 \cdot 4 & 5 \cdot 6 + 4 \cdot (0) \end{pmatrix} = \begin{pmatrix} 33 & -18 \\ 21 & 30 \end{pmatrix}$$

An equivalent method of matrix-matrix multiplication is as follows: in the resulting matrix, the element in the i-th row and j-th column is equal to the dot product of the i-th row of the first matrix and the j-th column of the second matrix:

$$c_{ij} = \vec{A}_i \cdot \vec{B}^j$$

where here the lower index i means the i-th row vector, and an upper index j means the j-th column vector.

Write here more about rhe "element-wise" multiplication scheme

An important characteristic of matrix-matrix multiplication is that it is **not commutative**: $A \cdot B \neq B \cdot A$ (even if the matrices are both square matrices of equal dimensions).

Example 86

For $A = \begin{pmatrix} 2 & 7 \\ 1 & 3 \end{pmatrix}$, $B = \begin{pmatrix} 5 & 1 \\ 4 & 4 \end{pmatrix}$ - on one hand:

$$\overbrace{\begin{pmatrix} 2 & 7 \\ 1 & 3 \end{pmatrix}}^{A} \overbrace{\begin{pmatrix} 5 & 1 \\ 4 & 4 \end{pmatrix}}^{B} = \overbrace{\begin{pmatrix} 38 & 30 \\ 17 & 13 \end{pmatrix}}^{A \cdot B},$$

but on the other hand:

$$\overbrace{\begin{pmatrix} 5 & 1 \\ 4 & 4 \end{pmatrix}}^{B} \overbrace{\begin{pmatrix} 2 & 7 \\ 1 & 3 \end{pmatrix}}^{A} = \overbrace{\begin{pmatrix} 11 & 38 \\ 12 & 40 \end{pmatrix}}^{B \cdot A}.$$

We can see that $A \cdot B \neq B \cdot A$.

The identity matrix in matrix-matrix multiplication is the same one as in matrix-vector multiplication.

Example 87

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 4 \\ 7 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 7 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 7 & 2 \end{pmatrix}$$

The identity matrix has the property that for any matrix A,

$$I \cdot A = A \cdot I = A.$$

A repeated multiplication of a matrix A with itself n times is denoted as A^n.

Example 88

For $A = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$,

$$A^2 = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 8 \\ 0 & 9 \end{pmatrix}$$

$$A^3 = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 26 \\ 0 & 27 \end{pmatrix}$$

$$A^4 = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 80 \\ 0 & 81 \end{pmatrix}$$

$$\vdots$$

4.10 Matrix-Matrix Multiplication as Transformation Composition

Multiplying matrices from the left generates a matrix that represents a composition of the corresponding linear transformations.

Example 89

The matrix $A = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}$ rotates the space by 30° counter clock-wise. The matrix $B = \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}$ shears space by a factor of 3 in the x-axis. The product $A \cdot B$ of these two matrix is

$$\begin{aligned} C &= A \cdot B \\ &= \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{3\sqrt{3}-1}{2} \\ \frac{1}{2} & \frac{3+\sqrt{3}}{2} \end{pmatrix}. \end{aligned}$$

This product represents a linear transformation that **first** skews space by 3 in the x-axis, and **then** rotates it by 30° counter clock-wise.

The other product, $B \cdot A = \begin{pmatrix} \frac{3+\sqrt{3}}{2} & \frac{3\sqrt{3}-1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}$ represents a transformation that **first** rotates space by 30° counter clock-wise, and **then** skews it by 3 in the x-axis.

4.11 Trace of a Matrix

The *trace* $\mathrm{tr}(A)$ of a square $N \times N$ matrix A is the sum of its diagonal elements:

$$\mathrm{tr}(A) = \sum_{i=1}^{N} a_{ii}.$$

Example 90

The trace of $A = \begin{pmatrix} 1 & 5 & 7 & -3 \\ 0 & 3 & 9 & 5 \\ -2 & 10 & 3 & 9 \\ 4 & 0 & 0 & 7 \end{pmatrix}$ is

$$\mathrm{tr}(A) = 1 + 3 + 3 + 7 = 14.$$

The trace of a matrix-matrix product does not depend on the order of multiplication, i.e.

$$\mathrm{tr}(A \cdot B) = \mathrm{tr}(B \cdot A),$$

since

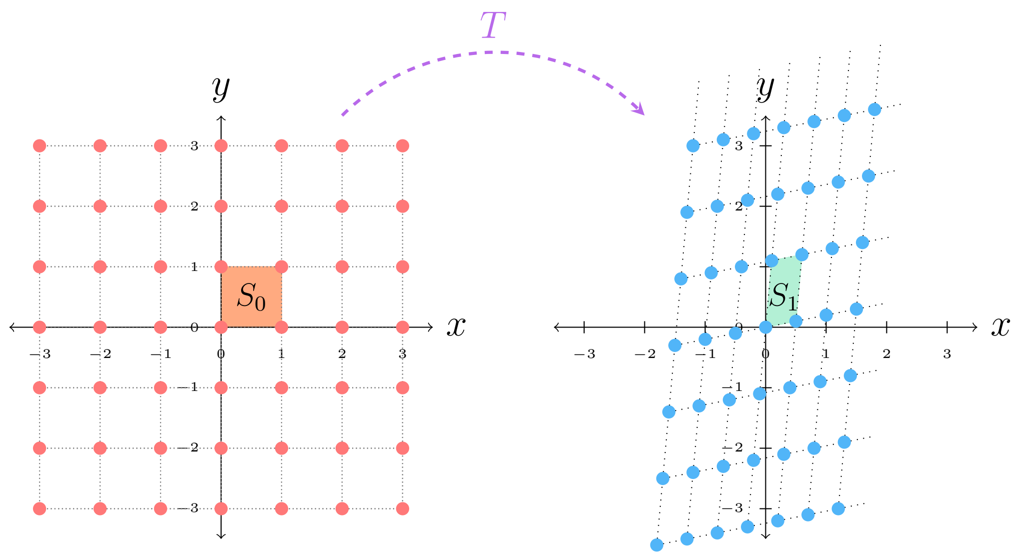
$$\mathrm{tr}(A \cdot B) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} b_{ji} = \mathrm{tr}(B \cdot A).$$

4.12 The Determinant

When viewing matrices as linear maps between vector spaces, the *determinant* is defined as the change in a unit area (or for higher dimensional spaces: the volume) before and after the transformation. Since linear transformations preserve ratios between areas, this change is uniform across the entire space.

Example 91

Applying the transformation represented by $A = \begin{pmatrix} 0.5 & 0.1 \\ 0.1 & 1.1 \end{pmatrix}$ to $\mathbb{R}^2$ yields the following:



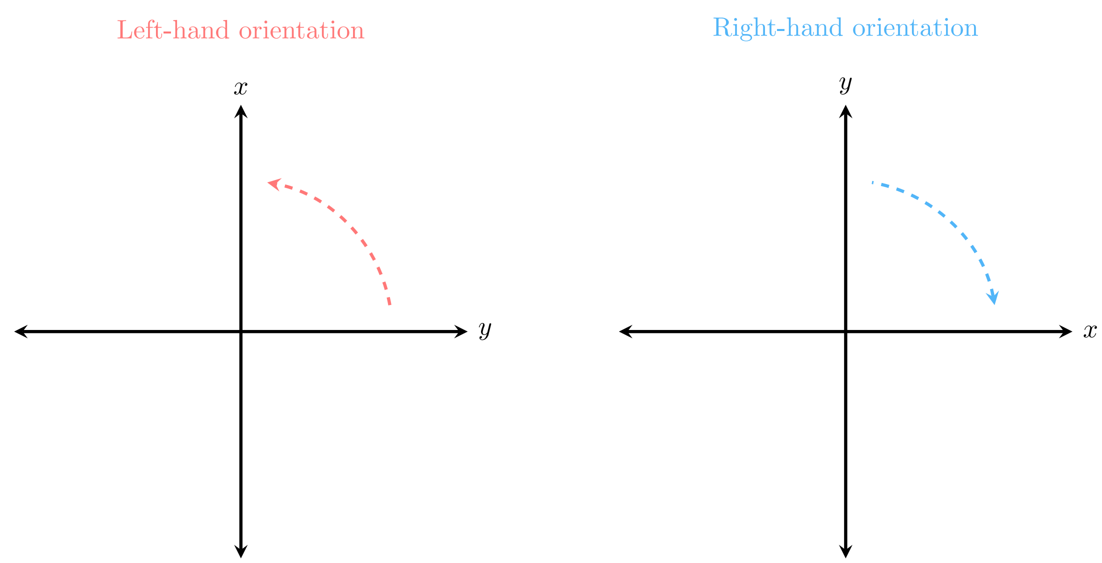
The area in orange on the left is $S_0 = 1$, and is transformed into the green area $S_1 = 0.54$ on the right.

The notation for the determinant of a matrix A is $\det(A)$ or $|A|$.

The determinant could be a positive value, a negative value or be equal to zero.

4.12.1 Negative Determinant

A negative determinant means that the transformation changes the orientation of the space (which can be determined by the right-hand rule). In $\mathbb{R}^2$ the normal orientation used (right-handed orientation) is the one where the x-axis is to the right (clock-wise) to the x-axis. The flipped orientation (left-handed orientation) is the opposite:

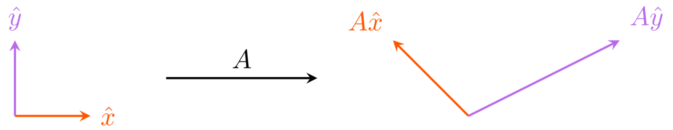


Example 92

The transformation represented by $A = \begin{pmatrix} -1 & 2 \\ 1 & 1 \end{pmatrix}$ flips the orientation of $\mathbb{R}^2$. This is evident by looking at how $\hat{x}$ and $\hat{y}$ transform, given by the columns of A, i.e.

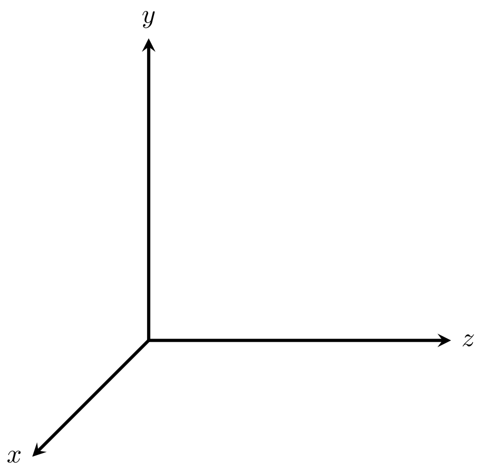
$$\hat{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \xrightarrow{A} \begin{pmatrix} -1 \\ 2 \end{pmatrix} = A\hat{x}$$
$$\hat{y} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \xrightarrow{A} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = A\hat{y}$$

Drawing $\hat{x}, \hat{y}$ vs. $\begin{pmatrix} -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ reveals that the orientation of space has flipped:

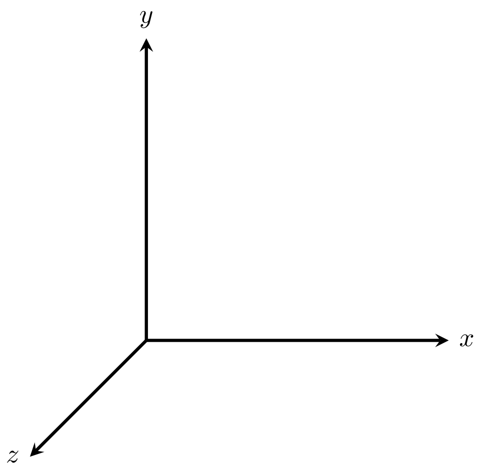


In $\mathbb{R}^3$ the orientations look as follows:

Left-hand orientation



Right-hand orientation



4.12.2 Zero Determinant

A zero determinant means that the dimensionality of the space after application of the matrix is lower than the dimensionality before the application. This can be easily understood in $\mathbb{R}^2$: a transformation that transforms all vectors to a single line results in a change of area equaling zero (since a line has no area), and thus the determinant of the matrix representing the transformation is also zero.

Example 93

The transformation represented by the matrix $A = \begin{pmatrix} 3 & 6 \\ -1 & -2 \end{pmatrix}$ transfomrs all vectors to the same line.

For example:

$$A\begin{pmatrix}-1\\2\end{pmatrix}=\begin{pmatrix}-3+12\\1-4\end{pmatrix}=\begin{pmatrix}9\\-3\end{pmatrix}$$
$$A\begin{pmatrix}0\\4\end{pmatrix}=\begin{pmatrix}0+24\\0-8\end{pmatrix}=\begin{pmatrix}24\\-8\end{pmatrix}$$
$$A\begin{pmatrix}2\\-7\end{pmatrix}=\begin{pmatrix}6-42\\-2+14\end{pmatrix}=\begin{pmatrix}-36\\12\end{pmatrix}$$

Notice how all these vectors are of the form $\begin{pmatrix}x\\-\frac{x}{3}\end{pmatrix}$, meaning that they all lie on the line $y=-\frac{1}{3}x$.

The determinant of a matrix is zero if its columns (or rows) are **linearly depended**: that is the case since linearly depended columns mean the the matrix transform the standard basis vectors to a set of vectors that are linearly depended, i.e. of a lower dimensionality.

Since a matrix with determinant zero transforms space to a lower dimensionality subspace, the transformation it represents is **not invertible**, i.e. the transformation is not a injective transformation.

4.12.3 Calculating the Determinant

The determinant of a 2×2 matrix $A=\begin{pmatrix}a&b\\c&d\end{pmatrix}$ is

$$|A|=ad-cb,$$

or graphically:

$$\begin{vmatrix}a&b\\c&d\end{vmatrix}.$$

Example 94

The determinant of $A=\begin{pmatrix}3&-1\\2&7\end{pmatrix}$ is

$$|A|=3\cdot 7-(-1)\cdot 2=21+2=23.$$

The determinant of $B=\begin{pmatrix}10&49\\100&-3\end{pmatrix}$ is

$$|B|=10\cdot(-3)-49\cdot 100=-30-4900=-4930.$$

The determinant of $C=\begin{pmatrix}21&7\\9&3\end{pmatrix}$ is

$$|C|=21\cdot 3-7\cdot 9=63-63=0.$$

The determinant of a 3×3 matrix is a bit more complicated. We first define the *ij-minor* (symbol: m_{ij}) of a 3×3 matrix A as the 2×2 matrix resulting when discarding the i-th row and j-th column of A.

Example 95

The matrix $A=\begin{pmatrix}1&3&9\\0&2&-1\\5&6&5\end{pmatrix}$ has 9 minors. The 11-minor can be found by hiding the 1st row and

1st column of A:

$$\begin{pmatrix} \text{[redacted]} & & \\ \text{[redacted]} & 2 & -1 \\ \text{[redacted]} & 6 & 5 \end{pmatrix} \longrightarrow m_{11} = \begin{pmatrix} 2 & -1 \\ 6 & 5 \end{pmatrix}.$$

Similarily, the 23-minor is

$$\begin{pmatrix} 1 & 3 & \text{[redacted]} \\ \text{[redacted]} & & \\ 5 & 6 & \text{[redacted]} \end{pmatrix} \longrightarrow m_{23} = \begin{pmatrix} 1 & 3 \\ 5 & 6 \end{pmatrix}.$$

Using the minors, the determinant of a 3×3 matrix $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$ is

$$|A| = a_{11}\,|m_{11}| - a_{12}\,|m_{12}| + a_{13}\,|m_{13}|\,.$$

Example 96

Let us calculate the determinant of $A = \begin{pmatrix} -3 & -9 & 2 \\ -1 & -5 & 4 \\ -6 & 5 & 1 \end{pmatrix}$. First, we calculate the minors:

$$m_{11} : \begin{pmatrix} \text{[redacted]} & & \\ \text{[redacted]} & -5 & 4 \\ \text{[redacted]} & 5 & 1 \end{pmatrix} \longrightarrow m_{11} = \begin{pmatrix} -5 & 4 \\ 5 & 1 \end{pmatrix}$$

$$m_{12} : \begin{pmatrix} \text{[redacted]} & & \\ -1 & \text{[redacted]} & 4 \\ -6 & \text{[redacted]} & 1 \end{pmatrix} \longrightarrow m_{12} = \begin{pmatrix} -1 & 4 \\ -6 & 1 \end{pmatrix}$$

$$m_{13} : \begin{pmatrix} \text{[redacted]} & & \\ -1 & -5 & \text{[redacted]} \\ -6 & 5 & \text{[redacted]} \end{pmatrix} \longrightarrow m_{12} = \begin{pmatrix} -1 & -5 \\ -6 & 5 \end{pmatrix}$$

Thus, the determinant of A is:

$$\begin{aligned} |A| &= -3m_{11} + 9m_{12} + 2m_{13} \\ &= -3 \cdot \begin{vmatrix} -5 & 4 \\ 5 & 1 \end{vmatrix} + 9 \cdot \begin{vmatrix} -1 & 4 \\ -6 & 1 \end{vmatrix} + 2 \cdot \begin{vmatrix} -1 & -5 \\ -6 & 5 \end{vmatrix} \\ &= -3\,(-5-20) + 9\,(-1+24) + 2\,(-5-30) \\ &= (-3)\cdot(-25) + 9 \cdot 23 + 2 \cdot (-35) \\ &= 75 + 207 - 70 \\ &= 212. \end{aligned}$$

4.12.4 Determinant of Matrix-Matrix Product

Since the multiplication of two matrices represents a composition of the two corresponding linear transformations, the determinant of the product equals the product of the determinants of the two matrices:

$$|A \cdot B| = |A||B|.$$

For a product of n matrices this becomes

$$\left|\prod_{i=1}^{n} A_i\right| = \prod_{i=1}^{n} |A_i| \,.$$

4.13 Inverse Transformations

4.13.1 Inverse Matrices

If for some matrix A, there exist a matrix B such that $A \cdot B = I$, we say that A is *invertible*, and that B is the *inverse* matrix of A, and we denote it by A^{-1}.

Example 97

The inverse of $A = \begin{pmatrix} 1 & 3 \\ 2 & 5 \end{pmatrix}$ is $A^{-1} = \begin{pmatrix} -5 & 3 \\ 2 & -1 \end{pmatrix}$:

$$\begin{pmatrix} 1 & 3 \\ 2 & 5 \end{pmatrix}\begin{pmatrix} -5 & 3 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} 1 \cdot (-5) + 3 \cdot 2 & 1 \cdot 3 + (-1) \cdot 3 \\ 2 \cdot (-5) + (5) \cdot 2 & 2 \cdot (3) + 5 \cdot (-1) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Note 22

It is not a straight-forward task to find the inverse of a given matrix (don't be fooled by the nice structure seen in the above example!).

Of course, the multiplication of A and A^{-1} is commutative, since both products are equal to I.

As a linear mapping, an inverse of a matrix represents the inverse mapping: if $A\vec{v} = \vec{u}$, then $A^{-1}\vec{u} = \vec{v}$.

Example 98

Inverse transformation example

If a matrix does not have an inverse, it is called a *singular matrix*. Such matrices have a determinant equal to zero. The reason for this is that if a matrix has a determinant of zero, it transforms some space to a subspace of lower dimensionality. Thus, the transformation results in a lost of information, and can't be inverted.

Example 99

plane to line transformation?

4.13.2 Kernel and Nullspace of a Transformation

The *kernel* of a linear transformation T is the set of vectors that are transformed to the zero vector. When the transformation is represented by a matrix, we call the kernel of T the *nullspace* of A.

Example 100

The linear transformation represented by the matrix $A = \begin{pmatrix} 1 & 2 & -5 \\ -2 & -4 & 10 \\ 0.5 & 1 & -2.5 \end{pmatrix}$ affects the vectors $\vec{u} = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$ and $\vec{v} = \begin{pmatrix} 5 \\ 0 \\ 1 \end{pmatrix}$ as follows:

$$\begin{aligned} A\vec{u} &= \begin{pmatrix} 1 & 2 & -5 \\ -2 & -4 & 10 \\ 0.5 & 1 & -2.5 \end{pmatrix} \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \cdot (-2) + 2 \cdot 1 - 5 \cdot 0 \\ -2 \cdot (-2) - 4 \cdot 1 + 10 \cdot 0 \\ 0.5 \cdot (-2) + 1 \cdot 1 - 2.5 \cdot 0 \end{pmatrix} \\ &= \begin{pmatrix} -2+2 \\ 4-4 \\ 1-1 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \end{aligned}$$

$$\begin{aligned} A\vec{v} &= \begin{pmatrix} 1 & 2 & -5 \\ -2 & -4 & 10 \\ 0.5 & 1 & -2.5 \end{pmatrix} \begin{pmatrix} 5 \\ 0 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 \cdot 5 + 2 \cdot 0 + 5 \cdot 1 \\ -2 \cdot 5 - 4 \cdot 0 + 10 \cdot 1 \\ 0.5 \cdot 5 + 1 \cdot 0 - 2.5 \cdot 1 \end{pmatrix} \\ &= \begin{pmatrix} -5+5 \\ -10+10 \\ 2.5-2.5 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \end{aligned}$$

Thus, the vectors $\vec{u}, \vec{v}$ are in the kernel of T and the nullspace of A.

Any linear combination of the vectors in a kernel of a linear transformation T is also in the kernel of T:

$$\begin{aligned} T(\vec{u}) = T(\vec{v}) &= \vec{0} \\ &\Downarrow \\ T(\alpha\vec{u} + \beta\vec{v}) &= \alpha T(\vec{u}) + \beta T(\vec{v}) \\ &= \alpha\vec{0} + \beta\vec{0} \\ &= \vec{0}. \end{aligned}$$

Example 101

Based on the previous example, the vector $\vec{w} = 2\vec{u} - 3\vec{v} = \begin{pmatrix} -19 \\ 2 \\ -3 \end{pmatrix}$ is transformed by A as follows:

$$
\begin{aligned}
A\vec{w} &= \begin{pmatrix} 1 & 2 & -5 \\ -2 & -4 & 10 \\ 0.5 & 1 & -2.5 \end{pmatrix} \begin{pmatrix} -19 \\ 2 \\ -3 \end{pmatrix} \\
&= \begin{pmatrix} 1 \cdot (-19) + 2 \cdot 2 - 5 \cdot (-3) \\ -2 \cdot (-19) - 4 \cdot 2 + 10 \cdot (-3) \\ 0.5 \cdot (-19) + 1 \cdot 2 - 2.5 \cdot (-3) \end{pmatrix} \\
&= \begin{pmatrix} -19 + 4 + 15 \\ 38 - 8 - 30 \\ -9.5 + 2 + 7.5 \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.
\end{aligned}
$$

Thus, $\vec{w} = 2\vec{u} - 3\vec{v}$ is also in the nullspace of A.

If the kernel of a linear transformation is only the zero vector, then the transformation is invertible, and correspondingly has a determinant different than zero. If the kernel of a linear transformation has any other vector, than the transformation is not invertible, and its determinant equals zero.

The *rank* of a square matrix is the difference between its dimension N and the dimension M of its nullspace:

$$R = M - N.$$

Example 102

The matrix A from the previous two examples has a rank

$$R = 3 - 2 = 1.$$

Systems of Linear Equations 5

5.1 Basics

A *linear equation* is an equation that is a linear combination of variables and an additional free coefficient.

Example 103

The equation

$$7x + y - 3z = -4$$

is a linear equation.

A *system of linear equations* is a set of several linear equations with the same variables.

Example 104

The following is a system of linear equations of the variables x, y and z:

$$\begin{aligned} 2x - 7y + z &= 26 \\ -3x + y &= -9 \\ 9y - 4z &= -31 \end{aligned}$$

Notice that for the second equation the coefficient of z is zero, while in the last equation the coefficient of x is zero.

A system of linear equations can be written a matrix form as $A\vec{u} = \vec{v}$, where A is a matrix in which each element a_{ij} is the coefficient of the j-th variable in the i-th row, $\vec{u}$ is a column vector of the variables, and $\vec{v}$ is a column vector of the free variables.

Example 105

The previous system can be written in a matrix form as

$$\begin{pmatrix} 2 & -7 & 1 \\ -3 & 1 & 0 \\ 0 & 9 & -4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 26 \\ -9 \\ -31 \end{pmatrix}.$$

A general system of m linear equations in n variables $x_1, x_2, \dots, x_n$ can be written as

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$

$$
\begin{aligned}
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
&\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m,
\end{aligned}
$$

which in a matrix form is simply

$$
\begin{pmatrix}
a_{11} & a_{12} & \dots & a_{1n} \\
a_{21} & a_{22} & \dots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & \dots & a_{mn}
\end{pmatrix}
\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}
=
\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.
$$

5.2 Solution Set

A *solution* is an ordered set of values which correspond to the variables of the system, such that all of its equations are satisfied.

Example 106

The only solution for the previous system is

$$
\begin{aligned}
x &= 2 \\
y &= -3 \\
z &= 1,
\end{aligned}
$$

which in vector form can be written as

$$
\vec{u} = \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix}.
$$

Generally, a linear system might have any of the following:

- An **infinite** amount of distinct solutions.
- Only **a single** solution.
- **No solutions**.

The number of solutions depends on the properties of the system, which we will briefly explore in this chapter.

5.2.1 Geometric Interpretation of the Solution Set

Generally, A linear equation in two variables represents a line in $\mathbb{R}^2$, a linear equation in three variables represents a plane in $\mathbb{R}^3$, and so forth.

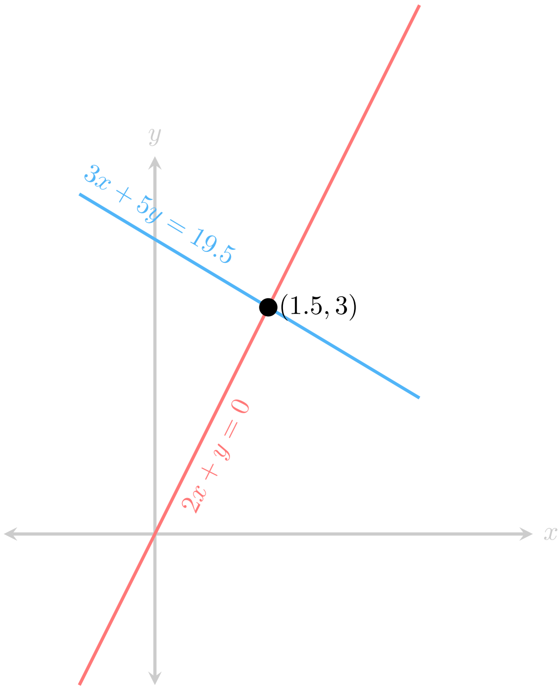
Thus, a solution of several linear equations represents a set of points where the respective shapes intersect.

Example 107

The equations

$$2x + y = 0$$
$$3x + 5y = 19.5$$

represent the following two lines:



As can be seen, the two lines intersect at $(1.5, 3)$. Thus, $x = 1.5, y = 3$ is a solution to the linear system represented by the two equations above.

The example above illustrates how two linear equations in two variables can have zero, one or infinite solutions:

- If the two equations represent two different **parallel** lines, then there is no intersection point between the lines and thus no solution to the system.
- If the two equations represent two lines with **different slopes** (as in the above example) then there is a single intersection point for the two lines and thus a single solution to the system.
- If the two equations represent **the same line**, then there is an infinite amount of intersection points and thus an infinite amount of solutions to the system.

For a system of equations with three variables, as each equation represents a plane in $\mathbb{R}^3$, any two of the equations can either intersect (the intersection being a line), be parallel or be the same plane.

And in the case of three equations in three variables, there can either be a single soultion (a point where all three planes intersect), no solution or infinite solutions.

5.2.2 General Number of Solutions

The number of equations and number of variables control the possible properties of the solution.

Example 108

A geometric representation of the number of possible solutions for systems in two variables:

5.3 Solving a System of Linear Equations

5.3.1 Eliminating Variables

5.3.2 Row Reduction (Gaussian Elimination)

As mentioned earlier, a system of linear equations can be written in a matrix form as

$$A\vec{u} = \vec{v},$$

where A represents the coefficients of the variables in their respective equation, $\vec{u}$ represents the variables $x_1, x_2, \dots, x_n$ and $\vec{v}$ represents the free coefficients $b_1, b_2, \dots, b_m$. In the Gaussian elimination scheme we "stick" A and $\vec{v}$ together, forming an augmented matrix of the following structure:

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right)$$

Example 109

The following system

$$\begin{aligned} -x + 3z &= 20 \\ 3x + y + 3z &= 15 \\ 9x + 3y &= -18 \end{aligned}$$

is written in the augmented matrix form as

$$\left(\begin{array}{ccc|c} -1 & 0 & 3 & 20 \\ 3 & 1 & 3 & 15 \\ 9 & 3 & 0 & -18 \end{array}\right).$$

Then, a series of steps are performed, where each step is one of the following type (called *elementry row operations*):

- Swapping two rows (notation: $R_i \leftrightarrow R_j$).
- Multiplying a row by a scalar (notation: $\alpha R_i \to R_i$).
- Adding to a row a scalar multiple of another row (notation: $R_i + \alpha R_j \to R_i$).

Example 110

Performing some elementry row operations on the previous matrix:

$$\left(\begin{array}{ccc|c} -1 & 0 & 3 & 20 \\ 3 & 1 & 3 & 15 \\ 9 & 3 & 0 & -18 \end{array}\right) \xrightarrow{-\frac{1}{R_1}\to R_1} \left(\begin{array}{ccc|c} 1 & 0 & -3 & -20 \\ 3 & 1 & 3 & 15 \\ 9 & 3 & 0 & -18 \end{array}\right) \xrightarrow{R_2-3R_1\to R_2}$$
$$\left(\begin{array}{ccc|c} 1 & 0 & -3 & -20 \\ 0 & 1 & 12 & 75 \\ 9 & 3 & 0 & -18 \end{array}\right) \xrightarrow{R_3-9R_1\to R_3} \left(\begin{array}{ccc|c} 1 & 0 & -3 & -20 \\ 0 & 1 & 12 & 75 \\ 0 & 3 & 27 & 162 \end{array}\right) \xrightarrow{R_3-3R_2\to R_3}$$
$$\left(\begin{array}{ccc|c} 1 & 0 & -3 & -20 \\ 0 & 1 & 12 & 75 \\ 0 & 0 & -9 & -63 \end{array}\right) \xrightarrow{-\frac{1}{9}R_3\to R_3} \left(\begin{array}{ccc|c} 1 & 0 & -3 & -20 \\ 0 & 1 & 12 & 75 \\ 0 & 0 & 1 & 7 \end{array}\right) \xrightarrow{R_1+3R_3\to R_1}$$
$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 12 & 75 \\ 0 & 0 & 1 & 7 \end{array}\right) \xrightarrow{R_2-12R_3\to R_2} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -9 \\ 0 & 0 & 1 & 7 \end{array}\right).$$

The process stops when we reach a form known as the *reduced row echelon form* of the matrix. This form has the following properties:

- The leading element in each non-zero row is equal to 1 (and is called a *leading 1*).
- In each row all elements before the leading 1 are equal to 0.
- The order of rows is such that the position of the leading 1s **increases** as we go down the rows.

Example 111

The following matrices are in the reduced row echelon forms:

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 5 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 4 & 2 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -9 \\ 0 & 0 & 1 & 7 \end{pmatrix}$$

The result of applying Gaussian elimination to the augmented matrix is a much simpler system to solve. In the example, the resulting matrix

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -9 \\ 0 & 0 & 1 & 7 \end{array}\right)$$

yields the solution directly:

$$x = 1,\ y = -9,\ z = 7.$$

However, the system in its row echelon form is not always as simple, and may require some more work isolating each variable. In addition, sometimes the row echelon form contains one or more rows of zeros, which indicate that the system is underdetermined.

5.3.3 Underdetermined Row Echelon Matrix

A row echelon form with one or more rows of zeros indicate an underdetermined system: in essence, the system has more variables than equations. This is because the row echelon form and the process to yield it actually tell us whether the rows of the matrix, when seen as vectors, form a linearly independent set. A row of zeros (or several of them) indicate that the set of row vectors comprising the matrix is linearly dependent, and thus at least one of the equations is a linear combination of the other equations. This in turn means that the set of equations has redundent information, and that the practical number of equations os lower than first seems.

Eigenvectors and Eigenvalues 6

6.1 Definition

An *eigenvector* of a transformation is a (non-zero) vector that doesn't change its direction when transformed, but is only scaled by a scalar λ. The scalar λ is its corresponding *eigenvalue*.

Generally, an *Eigenvalue equation* looks as following:

$$T\left(\boxed{\vec{v}}\right) = \overset{\text{Eigenvalue}}{\boxed{\lambda}}\,\boxed{\vec{v}},$$

Eigenvector

and in the equivalent matrix-vector multiplication form, where $T(\vec{v}) = A\vec{v}$:

$$A\vec{v} = \lambda\vec{v}.$$

Example 112

The transformation represented by the matrix $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ has several eigen vectors:

- The vector $\vec{u} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ is affected by A as follows:

$$A\vec{u} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

  which means it is an eigenvector of A with eigenvalue $\lambda = 1$.

- Similarily, the vector $\vec{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is also an eigenvector of A:

$$A\vec{v} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

  which means that the respective eigenvalue is $\lambda = 3$.

If a vector $\vec{u}$ is an eigenvector of a transformation T, then any scaling of $\vec{u}$, i.e. $\vec{v} = \alpha\vec{u}$, is also an eigenvector of T, with the same eigenvalue. This is due to the scaling property of linear transformations:

$$T(\vec{v}) = T(\alpha\vec{u}) = \alpha T(\vec{u}) = \alpha\lambda\vec{u} = \lambda(\alpha\vec{u}) = \lambda\vec{v}.$$

This equality is due to the scaling property of linear transformations

Note 23

While any scale of an eigenvector is itself an eigenvector, this is not true for general linear combinations. For example, let's examine the linear transformation represented by the matrix $\begin{pmatrix} 5 & 0 \\ 0 & -2 \end{pmatrix}$: Two eigenvectors of this transformation are

$$\vec{u}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ \vec{u}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

whith corresponding eigenvalues $\lambda_1 = 5,\ \lambda_2 = -2$.

However, the vector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, which is a linear combination of $\vec{u}_1$ and $\vec{u}_2$ (with $\alpha = \beta = 1$), is not an eigenvector of the transformation:

$$\begin{pmatrix} 5 & 0 \\ 0 & -2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 5+0 \\ 0+(-2) \end{pmatrix} = \begin{pmatrix} 5 \\ -2 \end{pmatrix} \neq \lambda\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

6.2 Characteristic Polynomial

As stated before, the general eigenvalue equation in matrix form is

$$A\vec{v} = \lambda\vec{v},$$

where A is a square matrix.

Rearranging the equation we get

$$A\vec{v} - \lambda\vec{v} = \vec{0},$$

and similar to scalar equations, we can group the vectors $\vec{v}$ together. In the matrix-vector case, we do this by using the identity matrix I:

$$(A - \lambda I)\vec{v} = \vec{0}.$$

In order to find all the possible eigenvectors λ, we must assume that $\vec{v}$ is non-zero (otherwise we find only the zero vector). This means that the matrix $A - \lambda I$ has a non-zero nullspace, which means its determinant is zero. Thus, solving the following equation:

$$|A - \lambda I| = 0,$$

will give as the values $\lambda_1, \lambda_2, \dots$ that are eigenvalues of the matrix A.

The resulting determinant $|A - \lambda I|$ is a polynomial of order n in λ, $P_n(\lambda)$, due to the way determinants are calculated. It is called the *characteristic polynomial* of A.

Example 113

Let's look again at the matrix $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$:

$$\begin{aligned} |A - \lambda I| &= \left| \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| \\ &= \left| \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right| \\ &= \begin{vmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{vmatrix} \\ &= (1-\lambda)(1-\lambda) - 4 \\ &= 1 - 2\lambda + \lambda^2 - 4 \\ &= \lambda^2 - 2\lambda - 3. \end{aligned}$$

Therefore, we should solve the equation

$$\lambda^2 - 2\lambda - 3 = 0,$$

which yields

$$\begin{aligned} \lambda_{1,2} &= \frac{2 \pm \sqrt{12+4}}{2} \\ &= \frac{2 \pm 4}{2} \\ &= 1 \pm 2 \\ &= -1, 3. \end{aligned}$$

Indeed, these are the two eigenvalues we saw earlier.

The characteristic polynomial of a diagonal matrix $D = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{pmatrix}$ is

$$p(\lambda) = \prod_{i=1}^{n} (d_i - \lambda),$$

i.e. - its eigenvalues are the diagonal values $d_1, d_2, \dots, d_n$.

Example 114

The characteristic polynomial of the matrix $A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 7 \end{pmatrix}$ is

$$p(\lambda) = (1-\lambda)(3-\lambda)(3-\lambda)(7-\lambda),$$

and thus has the eigenvalues $\lambda_1 = 1, \lambda_2 = 3, \lambda_3 = 3$ and $\lambda_4 = 7$.
(the respective eigenvectors are the basis vectors $\hat{e}_1, \hat{e}_2, \hat{e}_3$ and $\hat{e}_4$.)

6.3 Eigenspaces, Multiplicity

The space spanned by the eigenvectors of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ (or its square matrix representation, A) can have dimensionality smaller than or equal to n.

Example 115

The matrix $A = \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}$ has a single eigenvector $\vec{v} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, with eigenvalue $\lambda = 1$. Thus, it spans a single-dimensional space.

The matrix $B = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$ has two eigenvectors:

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \vec{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

with $\lambda_1 = 4, \lambda_2 = -2$, respectively. These eigenvectors span a 2-dimensional space.

The matrix $C = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is a rotation matrix (by $°90$ ccw), and thus has no eigenvectors.

Some matrices have repeating eigenvectors (i.e. $\lambda_i = \lambda_j$ for some indices i, j).

Example 116

The characteristic polynomial of the matrix $A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ is

$$p(\lambda) = (2 - \lambda)(-5 - \lambda)(2 - \lambda),$$

and thus its eigenvalues are $\lambda_1 = 2, \lambda_2 = -5$ and $\lambda_3 = 2$. We can see that $\lambda_1 = \lambda_3 = 2$.

(the respective eigenvectors are the basis vectors $\hat{e}_1$, $\hat{e}_2$ and $\hat{e}_3$.)

The number of repetition of an eigenvalue is called its *algebraic multiplicity*.

Example 117

In the previous example, the algebraic multiplicity of $\lambda = 2$ is 2, while the algebraic multiplicity of $\lambda = -5$ is 1.

The dimensionality of the space spanned by eigenvectors of the same eigenvalue is called the *geometric multiplicity* of the eigenvalue.

Example 118

In the previous example, the two eigenvectors that correspond to the eigenvalue $\lambda = 2$ are $\vec{u} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and $\vec{v} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ (i.e. $\hat{e}_1$ and $\hat{e}_3$). Together they span a 2-dimensional subspace of $\mathbb{R}^3$. Therefore, the geometric multiplication of $\lambda = 2$ is equal to 2.

The geomtric multiplication of an eigenvalue can be **at most** equal to its algebraic multiplication.

6.4 Matrix Diagonalization

A square matrix A is called *diagonalizable* if there exists an invertible matrix P (i.e. with non-zero determinant) such that

$$P^{-1}AP = D,$$

where D is a diagonal matrix. This can be alternatively formulated by multiplying both sides of the equation by P from the left and P^{-1} from the right, yielding

$$\underbrace{PP^{-1}}_{I}A\underbrace{PP^{-1}}_{I} = PDP^{-1}$$

$$\Downarrow$$

$$A = PDP^{-1}.$$

Example 119

The matrix $A = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 1 \\ -1 & 0 & 1 \end{pmatrix}$ can be written as PDP^{-1} for

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & -1 \end{pmatrix}, \quad D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We will show this by multiplying D with P^{-1} from the right and P from the left:

$$P^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ -1 & 0 & -1 \end{pmatrix},$$

and thus

$$\begin{aligned} DP^{-1} &= \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ -1 & 0 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 2 & 2 \\ 2 & 0 & 0 \\ -1 & 0 & -1 \end{pmatrix} \\ &\Downarrow \\ PDP^{-1} &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & -1 \end{pmatrix} \begin{pmatrix} 2 & 2 & 2 \\ 2 & 0 & 0 \\ -1 & 0 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 1 \\ -1 & 0 & 1 \end{pmatrix} \\ &= A. \end{aligned}$$

A matrix that is **not diagonalizable** is called a *defective matrix*.

WRITE: WHY IS DIAGONALIZATION IMPORTANT.

6.5 Summary of Eigenvectors and Eigenvalues in $\mathbb{R}^2$

Eigenvectors and eigenvalues in $\mathbb{R}^2$

	Scaling	Unequal scaling	Rotation
Matrix	$\begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix}$	$\begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix}$	$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$
Characteristic polynomial	$(\lambda - k)$	$(\lambda - k_1)(\lambda - k_2)$	$\lambda^2 - 2\cos(\theta)\lambda + 1$
Eigenvalues	$\lambda_1 = \lambda_2 = k$	$\lambda_1 = k_1,\ \lambda_2 = k_2$	-
Algebraic multiplication	$\mu = 2$	$\mu_1 = \mu_2 = 1$	-
Geometric multiplication	$\gamma = 2$	$\gamma_1 = \gamma_2 = 1$	-
Eigenvectors	All vectors	$\vec{u}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ \vec{u}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$	-
	x-shear	y-shear	Reflection
Matrix	$\begin{pmatrix} 1 & k_x \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ k_y & 1 \end{pmatrix}$	$\begin{pmatrix} \cos(2\theta) & -\sin(2\theta) \\ \sin(2\theta) & \cos(2\theta) \end{pmatrix}$
Characteristic polynomial	$(1 - \lambda)^2$	$(1 - \lambda)^2$	$\lambda^2 - 2\cos(2\theta)\lambda + 1$
Eigenvalues	$\lambda_1 = \lambda_2 = 1$	$\lambda_1 = \lambda_2 = 1$	-
Algebraic multiplication	$\mu = 2$	$\mu = 2$	-
Geometric multiplication	$\gamma = 1$	$\gamma = 1$	-
Eigenvectors	$\vec{u} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$	$\vec{u} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$	-

7 Real-Life Uses of Linear Algebra

7.1 Stable Populations in Predator-Pray Models / Leslie Models?

Linear algebra sometimes shows up in rather surprising circumstances.

For example, a population of lions and zebras live in the same area. Each year i, the number of lions L_i depends on the number of lions in the previous year L_{i-1}, but also on the number of zebras in that year, Z_{i-1} (as the lions need zebras for food). The dependency is

$$L_i = 0.86L_{i-1} + 0.08Z_{i-1}.$$

Similarly, the number of zebras Z_i depends on the number of zebras in the previous year (as zebras reproduce), but also on the number of lions in the previous year, since again - the lions eat the zebras:

$$Z_i = -0.12L_{i-1} + 1.14Z_{i-1}.$$

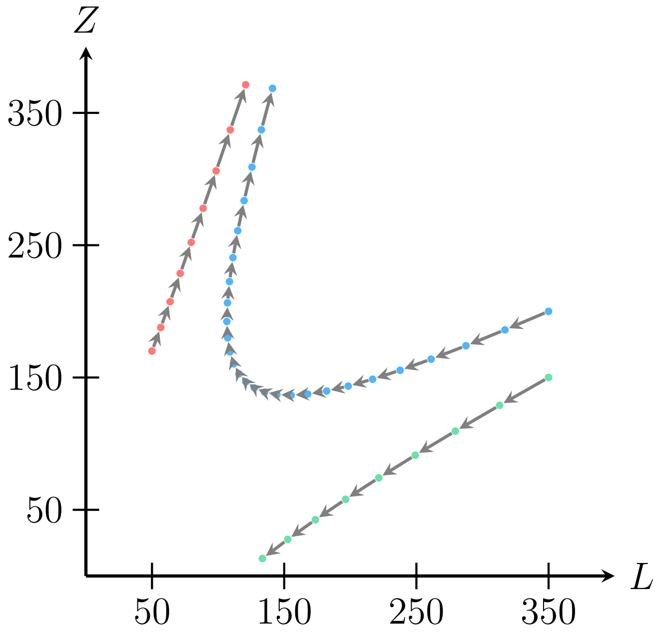
Together, the relation can be rewritten as (replacing i with $i+1$ and $i-1$ with i)

$$\begin{cases} L_{i+1} = & 0.86L_i + 0.08Z_{i-1} \\ Z_{i+1} = & -0.12L_i + 1.14Z_{i-1} \end{cases},$$

and can be expressed as a matrix-vector multiplication:

$$\begin{pmatrix} L_{i+1} \\ Z_{i+1} \end{pmatrix} = \begin{pmatrix} 0.86 & 0.08 \\ -0.12 & 1.14 \end{pmatrix} \begin{pmatrix} L_i \\ Z_i \end{pmatrix}.$$

We can graph the vector $\begin{pmatrix} L_i \\ Z_i \end{pmatrix}$ as a point in a coordinate system, and follow how the numbers of lions and zebras evolve every year, starting from $L = 350,\ Z = 150$ (red) and $L = 25,\ Z = 100$ (blue):

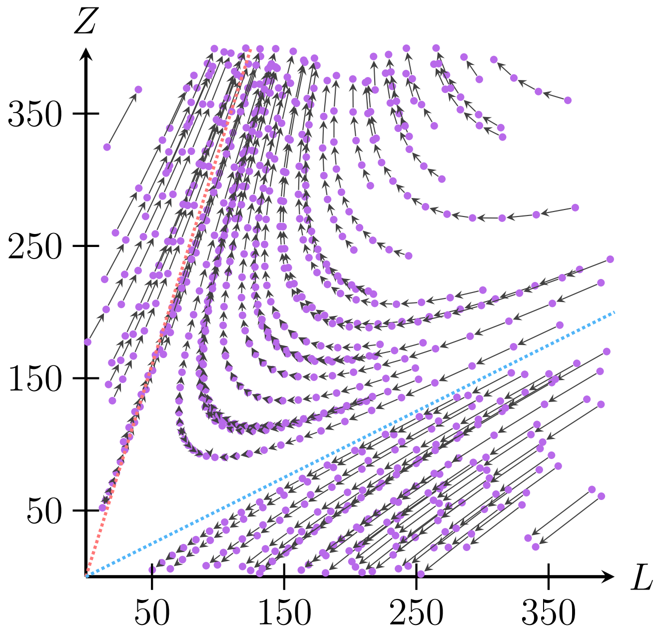


These three examples all behave differently from eachother: in the sequence drawn in red (starting at $\begin{pmatrix}50\\170\end{pmatrix}$) the numbers of both lions and zebras are increasing in each subsequent year.

In the sequence drawn in blue (starting at $\begin{pmatrix}350\\200\end{pmatrix}$) the numbers of both lions and zebras decline at first, but then at around 10th year, when there are about 105 lions and 140 zerbras, the number of zebras starts to increase rapidly, with the number of lions inreasing too starting from about the 13th year.

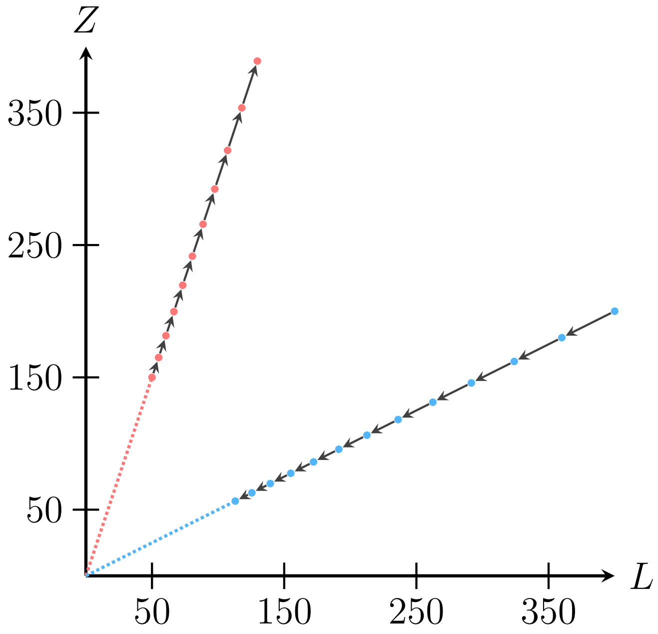
In the last sequence which is drawn in green (starting at $\begin{pmatrix}350\\150\end{pmatrix}$), the numbers of both lions and zebras decreases each year from the beginning.

Drawing many more such sequences reveals a pattern:



We can see two lines, here drawn in red and blue. Above the red line, the number of both lions and zebras always increses. Below the blue line, the number of both lions and zebras always decreases. Between the two lines, the number of both lions and zenras eventually increases, and depending on the inital position they might decrease at first.

What happens to points on these lines? Let's take a look:



As can be seen, after applying the tranformation to points on these lines, the results stay on the lines. This means that the lines represent eigenvectors of the tranformation.

Calculating the possible eigenvectors of the matrix $\begin{pmatrix} 0.86 & 0.08 \\ -0.12 & 1.14 \end{pmatrix}$ we get

$$\vec{u}_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix},\ \vec{u}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix},$$

with the corresponding eigenvalues being $\lambda_1 = 1.1, \lambda_2 = 0.9$. This means that points on the lines spanned by $\vec{u}_1$ and $\vec{u}_2$ (separetly) are scenarios where the ratio between the number of lions and of zebras stay stable over time.

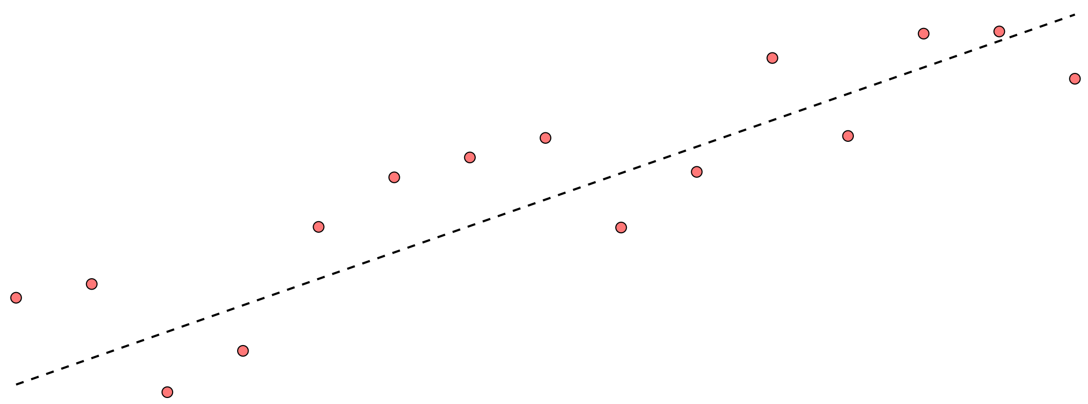
7.2 Least Squares Approximation

7.2.1 Linear Functions

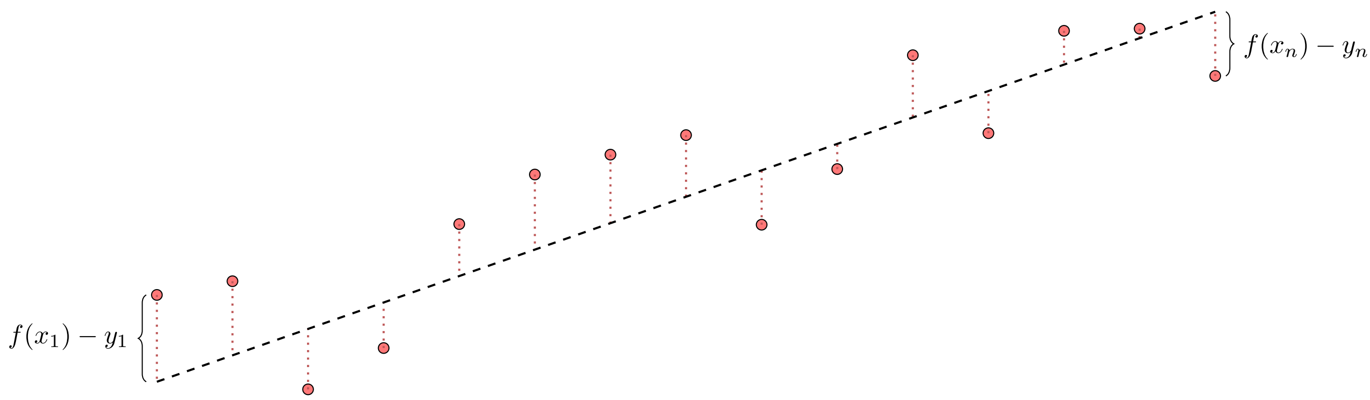
Say we have a set of n observed measurements stored as 2D points, i.e.

$$\begin{aligned} p_1 &= (x_1, y_1) \\ p_2 &= (x_2, y_2) \\ &\vdots \\ p_n &= (x_n, y_n) \end{aligned}$$

and we wish to find a line $f(x) = ax + b$ that best approximates these points, e.g.



We will consider such an approximation as the one for which the sum of the distances from each y_i to its expected value $f(x_i)$, i.e. $\sum_{i=1}^{n}\left[f(x_n)-y_n\right]$, is minimal.



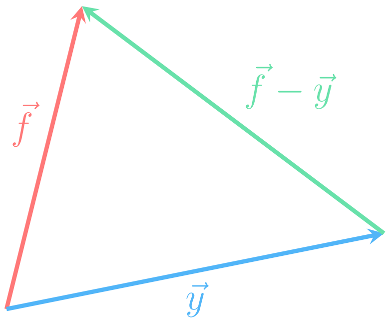
If we define the vectors $\vec{f}=\begin{pmatrix} f(x_1)\\ f(x_2)\\ \vdots\\ f(x_n)\end{pmatrix}$ and $\vec{y}=\begin{pmatrix} y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix}$, the sum of the distances $\sum_{i=1}^{n}\left[f(x_n)-y_n\right]$ is then simply

$$\sqrt{\vec{f}-\vec{y}},$$

but this distance is problematic since it is hard to calculate (square roots are computationally expensive to calculate), and there's a chance the sum is negative. Therefore, we instead calculate the square distance

$$\left(\sum_{i=1}^{n}\left[f(x_n)-y_n\right]\right)^2=\left(\vec{f}-\vec{y}\right)^2.$$

Therefore, for the best approximation we would want to find a and b such that $\left(\vec{f}-\vec{y}\right)^2$ is minimal. This of course will happen when $\vec{f}-\vec{y}$ is minimal. Drawing all these vectors helps to illustrate the problem geometrically:



The minimal distance is achieved when $\vec{f} - \vec{y}$ is orthogonal to $\vec{f}$.

Challange 12

Prove the last statement.

Of course, orthogonality happens when $\vec{f} \cdot \left(\vec{f} - \vec{y}\right) = 0$. We will now find the conditions that will enable this equality.

Explicitly, the vector $\vec{f}$ is

$$\vec{f} = \begin{pmatrix} ax_1 + b \\ ax_2 + b \\ \vdots \\ ax_n + b \end{pmatrix},$$

which can be written as a matrix A operated on a vector $\vec{v}$:

$$\vec{f} = A\vec{v} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}.$$

Designating the matrix as A and the vector as $\vec{v}$ yields $A\vec{v}$, and we wish to find the condition for which

$$A\vec{v}\left(A\vec{v} - \vec{y}\right) = 0.$$

Multiplying the equation by $A\vec{v}$ and rearranging the equation gives

$$A\vec{v} \cdot A\vec{v} = A\vec{v} \cdot \vec{y}.$$

Both sides now are dot products of vectors. We can consider each as being products of a column vector with a row vector, and write them as matrices (switching the dot product with a matrix-matrix product). Then, the matrices on the right must be transposed:

$$\left(A\vec{v}\right)^{\top} A\vec{v} = \left(A\vec{v}\right)^{\top} \vec{y}.$$

Due to the rules of transposing a matrix product ($\left(AB\right)^{\top} = B^{\top}A^{\top}$, see Equation xxx), we get (writing the column vector $\vec{v}$ as a matrix V)

$$V^{\top}A^{\top}A\vec{v} = V^{\top}A^{\top}\vec{y}.$$

Now, since $V^{\top}$ is simply a matrix with one row, we can remove it from both sides without changing the equality, yielding:

$$A^{\top}\vec{v} = A^{\top}vecy.$$

Thus, to find a and b that will yield the minimum total distance to of the line $f(x) = ax + b$ to the points $\{p_i\}$, we simply need to solve this linear system (which is easier than it looks!).

Example 120

Let's look at 6 points:

$$\begin{aligned}
p_1 &= (-2, -7.3)\\
p_2 &= (-1, -3.9)\\
p_3 &= (0, -1.2)\\
p_4 &= (1, 2.4)\\
p_5 &= (2, 4.7)\\
p_6 &= (3, 7.7)
\end{aligned}$$

The linear system we need to solve is thus

$$\begin{pmatrix} -2 & -1 & 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -2 & -1 & 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -7.3 \\ -3.9 \\ -1.2 \\ 2.4 \\ 4.7 \\ 7.7 \end{pmatrix}.$$

Multiplying both matrix-matrix products yields

$$\begin{pmatrix} 19 & 3 \\ 3 & 6 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 53.4 & 2.4 \end{pmatrix},$$

which when solved for a and b yields

$$a = 2.98 \quad b = -1.09.$$

7.2.2 General Polynomials

The method above can yield an approximation that is a general polynomial, which is an expressions of the form

$$P_m(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m = \sum_{k=0}^{m} a_k x^k,$$

with the restriction $n \geq m + 1$ (**note**: n is the number of points, while m is the degree of the polynomial!).

The difference in that case would be that the vectors $\vec{f}$ and $\vec{y}$ would be n-dimensional and the vectors $\vec{u}$ $(m+1)$-dimensional:

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \vec{v} = \begin{pmatrix} a_m \\ a_{m-1} \\ \vdots \\ a_1 \\ a_0 \end{pmatrix},$$

while the matrix A would be of dimension $n \times (m+1)$, and has the form

$$A = \begin{pmatrix} x_1^m & x_1^{m-1} & \dots & x_1 & 1 \\ x_2^m & x_2^{m-1} & \dots & x_2 & 1 \\ \vdots & & & & \\ x_n^m & x_n^{m-1} & \dots & x_n & 1 \end{pmatrix}.$$

Note 24

The reason that the dimension of $\vec{u}$ is $m+1$ and of A is $n \times (m+1)$, instead of m, is that we must include the free variable a_0 of the polynomial. This is reflected in both the vector $\vec{u}$ were it is seen explicitly, and in the matrix A where it is represented by the value 1.

Example 121

Using the same points from before, we can try and find a polynomial of order $n = 3$ that best approximates these points. This polynomial is of the form

$$P(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3.$$

The least square method would yield the following matrix and vectors:

$$A = \begin{pmatrix} (-2)^3 & (-2)^2 & -2 & 1 \\ (-1)^3 & (-1)^2 & -1 & 1 \\ (0)^3 & (0)^2 & 0 & 1 \\ (1)^3 & (1)^2 & 1 & 1 \\ (2)^3 & (2)^2 & 2 & 1 \\ (3)^3 & (3)^2 & 3 & 1 \end{pmatrix} = \begin{pmatrix} -8 & 4 & 2 & 1 \\ -1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 8 & 4 & 2 & 1 \\ 27 & 9 & 3 & 1 \end{pmatrix},$$

$$\vec{v} = \begin{pmatrix} a_3 \\ a_2 \\ a_1 \\ a_0 \end{pmatrix}, \ \vec{y} = \begin{pmatrix} -7.3 \\ -3.9 \\ -1.2 \\ 2.4 \\ 4.7 \\ 7.7 \end{pmatrix}.$$

As before, we calculate the products $A^\top A$ and $A^\top \vec{y}$, which yield the system

$$\begin{pmatrix} 859 & 243 & 115 & 27 \\ 243 & 115 & 27 & 19 \\ 115 & 27 & 19 & 3 \\ 27 & 19 & 3 & 6 \end{pmatrix} \begin{pmatrix} a_3 \\ a_2 \\ a_1 \\ a_0 \end{pmatrix} = \begin{pmatrix} 310.2 \\ 57.4 \\ 53.4 \\ 2.4 \end{pmatrix},$$

which when solved yields

$$a_3 = 0.004, \ a_2 = -0.07, \ a_1 = 3.0312, \ a_0 = -0.9111,$$

i.e. the polynomial

$$P(x) = 0.004x^3 - 0.07x^2 + 3.0312x - 0.9111.$$

Note 25

Approximating these points with a polynomial of degree $n = 3$ (or any order $n > 1$ for that matter) gives a bad approximation, as these points were generated from a line.

7.2.3 Other Functions

The method can be extended to any linear-like function. One example is a generic exponential function,

$$f(x) = ae^{bx},$$

which at first look doesn't seem linear, but can be made linear by using a logarithm:

$$\begin{aligned}\log\left[f(x)\right] &= \log\left(Be^{Ax}\right)\\ &= \log(B) + \log\left(e^{Ax}\right)\\ &= \log(B) + Ax.\end{aligned}$$

Thus, if we consider the vector $\vec{y} = \begin{pmatrix} A \\ \log(B) \end{pmatrix}$ and proceed with a linear approximation, we can use the least squares method for exponential functions as well.

Conversly, logarithmic functions of the type $f(x) = A\log(Bx)$ can be made linear by use of exponentiation.

7.3 Principal Component Analysis (PCA)

8 General Vector Spaces

Note: this chapter will be taught only if time permits.

In this chapter we will introduce a more general and abstract approach to vectors and linear transformations. While the approach is more rigorous in its mathematical approach than previous chapters, it is still by no means a formal text on the subject.

8.1 $\mathbb{R}^n$ as a Vector Space

Vectors have the follows properties (some of them may seem quite obvious, but are nontheless important as we will see soon):

- The addition of any two vectors is also a vector (closure under addition).

Example 122

For $\vec{u}=\begin{pmatrix}3\\-1\\0\end{pmatrix}, \vec{v}=\begin{pmatrix}4\\1\\2\end{pmatrix}$ (both in $\mathbb{R}^3$), the resulting vector $\vec{u}+\vec{v}=\begin{pmatrix}7\\0\\2\end{pmatrix}$ is also a vector in $\mathbb{R}^3$.

- For any two vectors $\vec{u},\vec{v}$: $\vec{u}+\vec{v}=\vec{v}+\vec{u}$ (i.e. vector addition is commutative).

Example 123

For the same vectors as above, $\vec{u}+\vec{v}=\vec{v}+\vec{u}=\begin{pmatrix}7\\0\\2\end{pmatrix}$.

- For any three vectors $\vec{u},\vec{v}$ and $\vec{w}$: $(\vec{u}+\vec{v})+\vec{w}=\vec{u}+(\vec{v}+\vec{w})$ (i.e. vector addition is associative).

Example 124

The vectors $\vec{u}=\begin{pmatrix}0\\1\end{pmatrix}, \vec{v}=\begin{pmatrix}3\\7\end{pmatrix}, \vec{w}=\begin{pmatrix}1\\-4\end{pmatrix}$ can be added as follows:

$$\left[\begin{pmatrix}0\\1\end{pmatrix}+\begin{pmatrix}3\\7\end{pmatrix}\right]+\begin{pmatrix}1\\-4\end{pmatrix}=\begin{pmatrix}3\\8\end{pmatrix}+\begin{pmatrix}1\\-4\end{pmatrix}=\begin{pmatrix}4\\4\end{pmatrix},$$

or as follows:

$$\begin{pmatrix}0\\1\end{pmatrix}+\left[\begin{pmatrix}3\\7\end{pmatrix}+\begin{pmatrix}1\\-4\end{pmatrix}\right]=\begin{pmatrix}0\\1\end{pmatrix}+\begin{pmatrix}4\\3\end{pmatrix}=\begin{pmatrix}4\\4\end{pmatrix},$$

which yields the same result.

- The zero vector $\vec{0}$ has the property that for any vector $\vec{v}$, $\vec{v}+\vec{0}=\vec{0}+\vec{v}=\vec{v}$.

Example 125

$$\begin{pmatrix}7\\1\\0\end{pmatrix}+\begin{pmatrix}0\\0\\0\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}+\begin{pmatrix}7\\1\\0\end{pmatrix}=\begin{pmatrix}7\\1\\0\end{pmatrix}.$$

- For any vector $\vec{u}$ there exist a vector $\vec{v}$ such that $\vec{u}+\vec{v}=\vec{v}+\vec{u}=\vec{0}$. Specifically, this is the vector $\vec{v}=-\vec{u}$.

Example 126

$$\begin{pmatrix}7\\1\\0\end{pmatrix}+\begin{pmatrix}-7\\-1\\0\end{pmatrix}=\begin{pmatrix}-7\\-1\\0\end{pmatrix}+\begin{pmatrix}7\\1\\0\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}.$$

In addition, the follows properties also exist together with scalars:

- For any vector $\vec{v}$ and any scalar α, the product $\alpha\cdot\vec{v}$ is also a vector (closure under multiplication).

Example 127

The $\mathbb{R}^4$ vector $\vec{v}=\begin{pmatrix}1\\-3\\2\\7\end{pmatrix}$ and the scalar $\alpha=-2$ multiplied together give

$$\alpha\vec{v}=-2\begin{pmatrix}1\\-3\\2\\7\end{pmatrix}=\begin{pmatrix}-2\\6\\-4\\-14\end{pmatrix},$$

which is also a vector in $\mathbb{R}^4$.

- For any two vectors $\vec{u},\vec{v}$ and any scalar α: $\alpha\cdot(\vec{u}+\vec{v})=\alpha\cdot\vec{u}+\alpha\cdot\vec{v}$.
- For any vector $\vec{v}$ and any scalars α,β: $(\alpha+\beta)\cdot\vec{v}=\alpha\cdot\vec{v}+\beta\cdot\vec{v}$.
- For any vector $\vec{v}$ and sny scalars α,β: $\alpha\cdot(\beta\cdot\vec{v})=(\alpha\cdot\beta)\cdot\vec{v}$.
- For any vector $\vec{v}$: $1\cdot\vec{v}=\vec{v}$.

Any set of elements obeying these properties, using any kind of operation between two of its elements (here the vector addition operation) and another operation between one element and a scalar (here scalar-vector multiplication) is called a *vector space*. The scalars must be of an algebraic structure called a *field*, which we will not define here and assume for simplicity that is either $\mathbb{R}$ or $\mathbb{C}$.

8.2 Examples of Vector Spaces

8.2.1 Polynomials of Degree $\leq$ n

A polynomial of a degree n is an expression of the form

$$\begin{aligned} P(x) &= a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \\ &= \sum_{k=0}^{n} a_k x^k, \end{aligned}$$

where $a_0, a_1, \ldots, a_n$ are called the *polynomial coefficients*.

When the value of a polynomial coefficient a_i is zero, we don't write the corresponding component.

Example 128

The 3rd degree polynomial with coefficients

$$a_0 = 2,\ a_1 = -5,\ a_2 = 0,\ a_3 = 1$$

is written explicitly as

$$p(x) = x^3 - 5x + 2.$$

(notice the descending power order used here, contrary to the ascending power order used in the definition)

Adding two polynimials together results in another polynomial, since the coefficients are added power-wise:

$$\begin{aligned} P(x) &= a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n, \\ Q(x) &= b_0 + b_1 x + b_2 x^2 + \cdots + b_n x^n, \\ \Rightarrow (P+Q)(x) &= (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots + (a_n + b_n)x^n. \end{aligned}$$

Multiplying a polynomial by a constant also results in a polynomial, where each of the powers are multiplied by the constant:

$$\alpha \cdot P(x) = \alpha a_0 + \alpha a_1 x + \alpha a_2 x^2 + \cdots + \alpha a_n x^n.$$

Example 129

Adding two polynomials:

$$\begin{aligned} P(x) &= 3x^2 + 2x - 5, \\ Q(x) &= 4x^2 + 7, \\ \Rightarrow (P+Q)(x) &= (3+4)x^2 + (2+0)x + (-5+7) = 7x^2 + 2x + 2. \end{aligned}$$

Multiplying a polynomial by a constant:

$$5 \cdot \left(x^3 - 5x^2 - x + 2\right) = 5x^3 - 25x^2 - 5x + 10.$$

We can use vector notation to represent a polynomial: that way, the i-th element of the vector corresponds to the coefficient a_i.

Example 130

Some polynomials represented as column vectors:

$$2x^5 - 7x^4 + 3x^2 - 1 \Rightarrow \begin{pmatrix} -1 \\ 0 \\ 3 \\ 0 \\ -7 \\ 2 \end{pmatrix}$$

$$x^2 - 2x + 1 \Rightarrow \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

$$-2x^7 + 5 \Rightarrow \begin{pmatrix} 5 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -2 \end{pmatrix}$$

It should be fairly obvious by now that polynomials of a degree $\leq n$ (with the addition and scalar multiplication shown here) constitute a vector space.

Challange 13

Prove that the polynomials of a degree $\leq n$ together with the shown addition and scalar multiplication are indeed a vector space.

Note 26

It is important to note that polynomials constitute a vector space only when we consider **all** the polynomials of a degree $\leq n$ in such a way that any polynomial of a lesser degree is considered to have the relevant higher order coefficients equaling zero. Otherwise, we lose the closure in regards to both polynomial addition and multiplication by a scalar.

The zero-element of the polynomial space is of course the zero-polynomial:
$P(x) = 0(a_0 = a_1 = a_2 = \cdots = a_n = 0)$. The inverse of any polynomial can be found by inverting each of its coefficients.

Since the polynomials are a well defined vector space, we can define an inner (dot) product for two polynomials $P(x) = \sum\limits_{k=0}^{n} a_k x^k$, $Q(x) = \sum\limits_{k=0}^{n} b_k x^k$:

$$\langle P, Q \rangle = a_0 b_0 + a_1 b_1 + a_2 b_2 + \dots a_n b_n = \sum_{k=0}^{n} (a_i + b_i) x^k.$$

Defining an inner product immidetly gives rise to a concept of orthogonality: if the inner product $\langle P, Q \rangle$ of two **non-zero** polynimials P, Q is zero, then they are orthogonal. This is usefull in many practical applications.

Example 131

All the polynomials $P^0(x) = 1, P^1(x) = x, P^2(x) = x^2, dots, P^n(x) = x^n$ are orthogonal, similar to how in the space $\mathbb{R}^n$ the standard basis vectors are orthogonal to each other. In fact, exactly as with $\mathbb{R}^n$, thes polynomials constitute an orthonormal basis of all polynomials with degree $\leq n$.

Since we can represent polynomials as column vectors, we can also use matrices to transform them. One such matrix is the follows:

$$D = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 2 & 0 & \dots & 0 \\ 0 & 0 & 0 & 3 & \dots & 0 \\ & & & \vdots & & \\ 0 & 0 & 0 & 0 & \dots & n \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix},$$

which represents non other than the differential (derivative) operator!

Example 132

Let's see the results of applying D to the polynimial $P(x) = 2x^3 + 5x^2 - 2x + 1$:

$$\begin{aligned} D\left[P\right] &= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \\ 5 \\ 2 \end{pmatrix} \\ &= \begin{pmatrix} \cancel{0 \cdot 1} + 1 \cdot (-2) + \cancel{0 \cdot 5} + \cancel{0 \cdot 2} \\ \cancel{0 \cdot 1} + \cancel{0 \cdot (-2)} + 2 \cdot 5 + \cancel{0 \cdot 2} \\ \cancel{0 \cdot 1} + \cancel{0 \cdot (-2)} + \cancel{0 \cdot 5} + 3 \cdot 2 \\ \cancel{0 \cdot 1} + \cancel{0 \cdot (-2)} + \cancel{0 \cdot 5} + \cancel{0 \cdot 2} \end{pmatrix} \\ &= \begin{pmatrix} -2 \\ 10 \\ 6 \\ 0 \end{pmatrix} \end{aligned}$$

The resulting represents the polynomial $Q(x) = 6x^2 + 10x - 2$, which is indeed the derivative of $P(x)$ (with respect to x).

Since the derivative operator can be represented by a matrix, it is a *linear operator* (at least for polynomials of degree $\leq n$).

Challange 14

Show that the derivative operator is a linear operator for any single variable real function.

8.2.2 Real Valued Functions

The set of real functions in the interval $[a, b]$ (where $a, b \in \mathbb{R}$, but could also be $-\infty$ and ∞, respectively) is a vector space together with function addition and multiplication by a real number:

- For any real functions f, g in $[a, b]$, the sum $f + g$ is also a function.

  Example 133

  The functions $f(x) = x^2 - 2e^x, g(x) = \sin(3x) - 4x^2$ can be added together to yield

  $$(f + g)(x) = -3x^2 - 2e^x + \sin(3x),$$

which is also a real function in $[a, b]$.

- Multiplying a real function f by a real scalar α also yields a real function.

Example 134

Multiplying the real function $f(x) = 5\log(x) - \frac{2}{x^3}$ by $\alpha = 7$ yields

$$(\alpha f)(x) = 35\log(x) - \frac{14}{x^3},$$

which is also a real function in $[a, b]$.

The rest of the criteria for a vector space are also fulfilled.

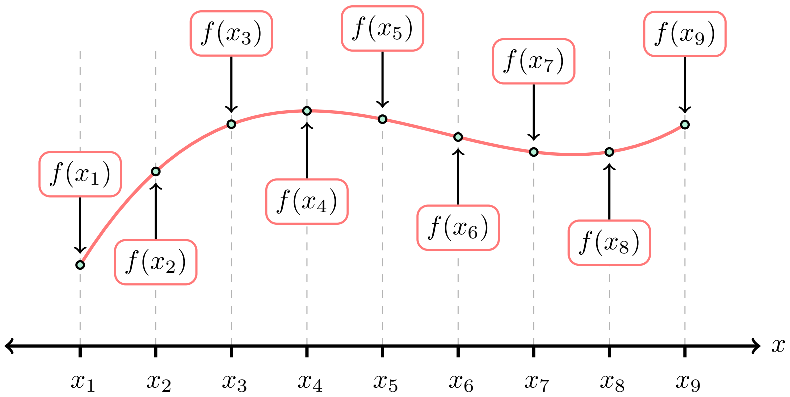
Challange 15

Show that the rest of the criteria are indeed fulfilled.

The dimensionality of this space is infinite. This can be shown by considering the interval $[a, b]$ to be discrete with n values, i.e.

$$\{x_1, x_2, \dots, x_n\},$$

where $x_1 = a, x_n = b$. Then, when discussing functions, we look at their values only in these discrete points:



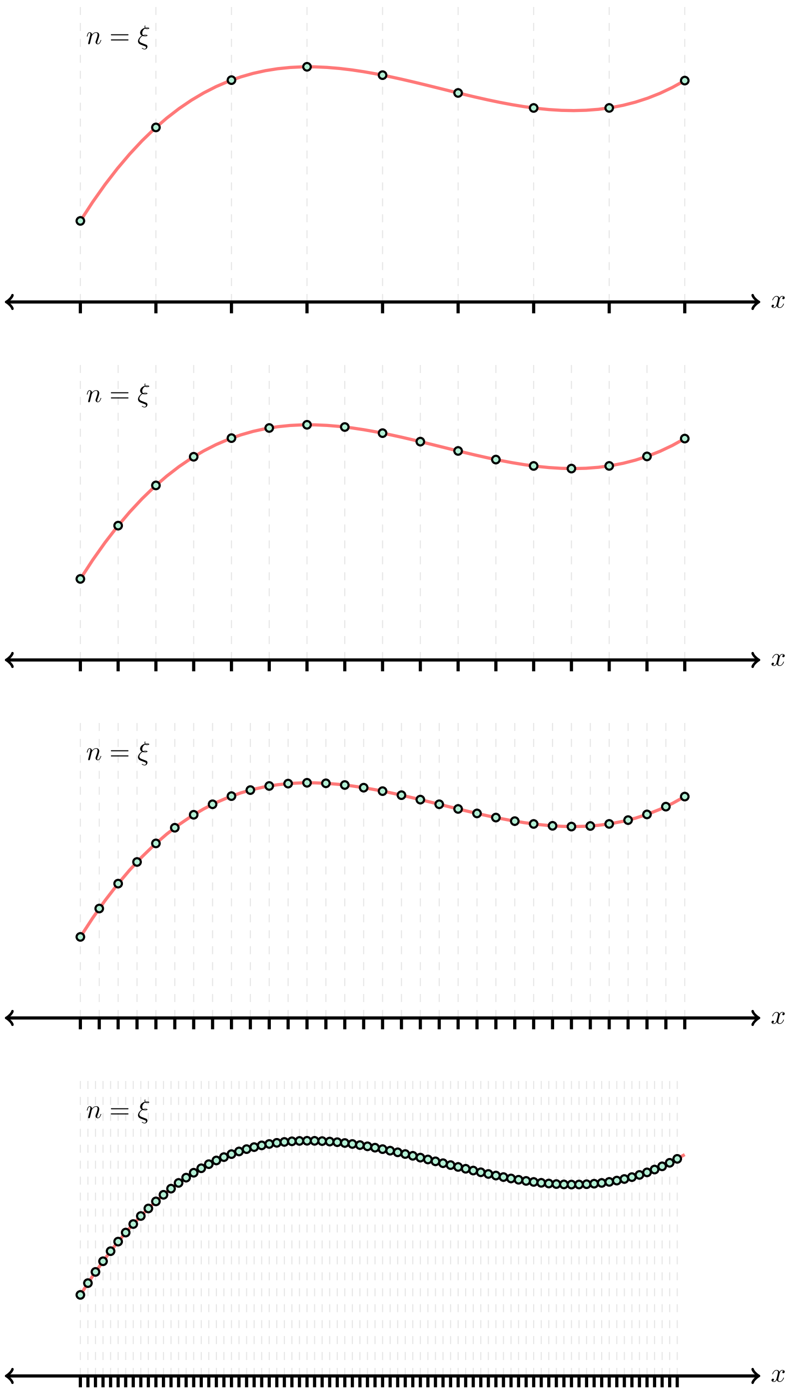
In this context, a function can be written as a vector with each component f_i having the value of the function in the respective x_i:

$$\boldsymbol{f} = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix}.$$

This sets implicitly a basis

$$B = \left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right\}.$$

Now we can increase the number of points n, and at infinity we will get the complete function:



Since each function has an infinite amount of components, the basis is also infinite, and is composed of dirac-detla "functions". The dirac delta is defined as:

$$\delta(x) = \begin{cases} +\infty & x = 0, \\ 0 & x \neq 0 \end{cases},$$

i.e. it is inifite at $x = 0$ and zero anywhere else. It can be visualized as so:

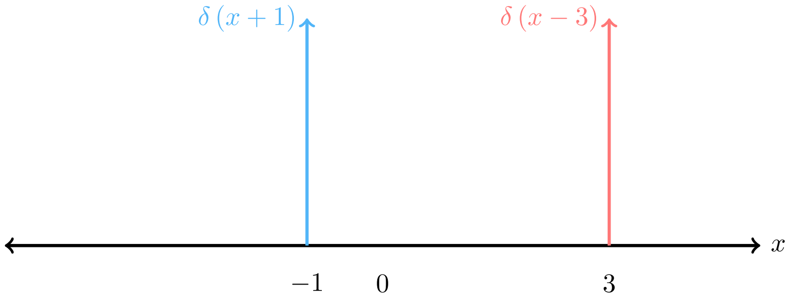
This means that the basis is composed of an infinite amount of dirac deltas of the the form

$$\delta\left(x-x'\right),$$

where x' is a number in $[a,b]$.

Example 135

A graphic representation of $\delta\left(x-3\right)$ and $\delta\left(x+1\right)$:



Spanning the Space

Inner Product and Normalization

We can also define an inner product on this space:

$$\langle f,g\rangle = \int_a^b f(x)g(x)\,\mathrm{d}x.$$

This leads to the idea of *orthogonal functions*.

Example 136

in the interval $[-\pi,\pi]$, the functions

$$f(x)=\sin(x),\;\; g(x)=x^2$$

have the follows inner product:

$$\begin{aligned}\langle f,g\rangle &= \int_{-\pi}^{\pi} f(x)g(x)\,\mathrm{d}x\\ &= \int_{-\pi}^{\pi} x^2\sin(x)\,\mathrm{d}x\end{aligned}$$

$$
\begin{aligned}
&= \int_{-\pi}^{0} x^2 \sin(x)\,\mathrm{d}x + \int_{0}^{\pi} x^2 \sin(x)\,\mathrm{d}x \\
&= \int_{-\pi}^{0} x^2 \sin(x)\,\mathrm{d}x - \int_{-\pi}^{0} x^2 \sin(x)\,\mathrm{d}x \\
&= 0.
\end{aligned}
$$

Thus, these two functions are orthogonal in $[-\pi, \pi]$.

Recall that in $\mathbb{R}^n$ the norm of a vector is defined as

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

which can be viewed as

$$\|\vec{v}\| = \sqrt{\sum_{i=1}^{n} (v_i \cdot v_i)}.$$

This is exactly the square root of the dot prouct of the vector with itself. Thus, a generalized norm for a vector $\boldsymbol{x}$ can be defined using the inner product as follows:

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}.$$

Thus, the norm of a real function $f(x)$ over $[a, b]$ is:

$$
\begin{aligned}
\|f\| &= \sqrt{\langle f, f \rangle} \\
&= \int_{a}^{b} f(x) f(x)\,\mathrm{d}x \\
&= \int_{a}^{b} f^2(x)\,\mathrm{d}x.
\end{aligned}
$$

Example 137

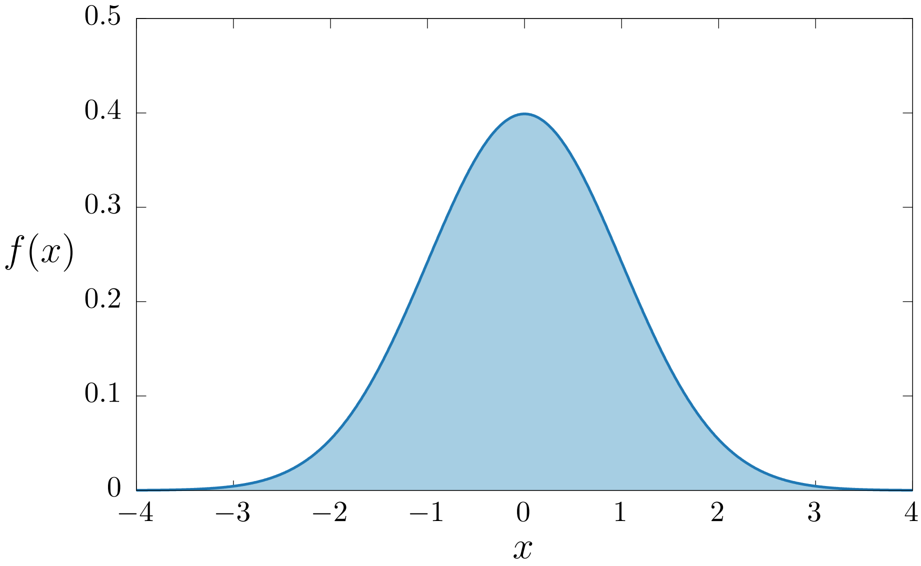
The norm of the Gaussian function $f(x) = e^{-\frac{x^2}{2}}$ over $(-\infty, \infty)$ is

$$
\begin{aligned}
\|f\| &= \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} e^{-\frac{x^2}{2}}\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} e^{-\frac{x^2}{2} - \frac{x^2}{2}}\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} e^{-x^2}\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}}.
\end{aligned}
$$

To normalize the function, we divide it by its norm, $\frac{1}{\sqrt{2\pi}}$, which yields the *Normal distribution*:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Graphing the normal distribution:



Eigenfunctions

Similar to the eigenvectors of a linear transformation, some linear operators operating on real functions have eigenfunctions: functions that are scaled by a real number when operated on by the operator.

Example 138

The derivative operator D has eigenfunctions. These are of course of the form

$$Df = \lambda f,$$

i.e. functions with a derivative that is equal to the function times a scalar. Their form is

$$f(x) = e^{\lambda x},$$

as the derivative of an exponential function is

$$\frac{\mathrm{d}f}{\mathrm{d}x} e^{\lambda x} = \lambda e^{\lambda x} = \lambda f(x).$$

Affine transformation, 55
Algebraic multiplicity of an eigenvalue, 88
Angle of a vector, 37
Anti-commutative, 46
Antidiagonal of a matrix, 59
Argument of a vector, 27, 37
Arguments of a function, 19

Basic math notation, 10
Basis, 41
Bijective function, 20
Binary trees, 26
Both, 9

Cardinality, 12
Cartesian coordinates, 37
Cartesian product, 16
Characteristic polynomial of a linear transformation, 86
Characteristic polynomial of a matrix, 86
Child node (graph theory), 25
Circle (graph theory), 24
Clique (graph theory), 26
Column vector, 32
Columns of a matrix, 58
Commutativity, 30
Complement of a set, 14
Complete graph, 26
Components of a vector, 32
Coordinates of a vector, 32
Counter diagonal of a matrix, 59
Cross product, 45
Cycle (graph theory), 24

Defective matrix, 89
Determinant, 70
Diagonal matrix, 60
Diagonalizable matrix, 89
Difference of a set, 14
Dimension of a vector, 35
Directed graph, 24
Direction of a vector, 27
Disjoint sets, 13
Distance (graph theory), 25
Domain of a function, 19
Dot product, 43

Edges, 23
Eigenvalue, 85
Eigenvalue equation, 85
Eigenvector, 85
Element-wise, 34
Elementry row operation, 82
Elements (of a set), 10
Elements of a matrix, 58

Empty set, 12

Field, 100
Function, 18

Geometric multiplicity of an eigenvalue, 88
Graphs, 23

Image of a function, 19
Injective function, 19
Inner nodes (graph theory), 25
Inner product, 43
Intersection, 12
Inverse of a matrix, 75
Invertible matrix, 75

Kernel of a matrix, 76
Kronecker delta, 64

Leading 1, 83
Length of a vector, 27, 36
Linear combination, 38
Linear equation, 79
Linear operator, 103
Linear transformation, 49
Linearly dependent, 38
Lower triangular matrix, 60

Magnitude of a vector, 27
Main diagonal of a matrix, 59
Major diagonal of a matrix, 59
Matrix, matrices, 58
Minor, 73
Minor diagonal of a matrix, 59

Nodes, 23
Norm of a vector, 27, 36
Normal distribution, 107
Normalized vector, 36
Nullspace, 76

Operator, 9
Origin of a vector space, 27, 28
Orthogonal, 42
Orthogonal basis, 42
Orthogonal functions, 106
Orthonormal basis, 42

Parallelogram, 30
Path (graph theory), 24
Perpendicular, 42
Polar coordinates, 37
Polynomial coefficients, 101
Powerset, 15
Primary diagonal of a matrix, 59
Principal diagonal of a matrix, 59
Projection of a vector, 44

Rank of a matrix, 77
Reduced row echelon form, 83
Relation between two sets, 17
Right-hand rule, 46
Root (graph theory), 25
Rooted trees (graph theory), 25
Row vector, 33
Rows of a matrix, 58

Scalar, 28
Scalar product, 43
Secondary diagonal of a matrix, 59
Shear matrix, 66
Singular matrix, 75
Skew matrix, 66
Solution of a linear system, 80
Square matrix, 59
Standard basis, 42
Subset, 12
Subspace, 41
Superset, 12
Surjective function, 20
System of linear equations, 79

Trace of a matrix, 70
Translation (transformation), 55
Transpose of a matrix, 60
Tree (graph theory), 25
Tree leaf, 25
Truth table, 10

Union of sets, 13
Upper triangular matrix, 60

Vector, 27
Vector space, 100
Venn diagram, 12
Vertex, vertices, 23

Zero vector, 30