Master in Bioinformatics for Health Sciences
(UPF-UB)

2007/08

Elements of Mathematics

Lecture Notes

Lecturer: Elisenda Feliu

1 Matrices and systems of linear equations

In this section we review the basic operations with matrices, and the discussion of systems of linear equations.

1.1 Basic concepts on matrices

Let K be a field. For simplicity, think of K as being the field of rational numbers $\mathbb{Q}$, the field of real numbers $\mathbb{R}$ or the field of complex numbers $\mathbb{C}$.

Definition 1.1. A matrix A with coefficients in the field K of order $n \times m$ is a collection of nm elements of K, indexed in the following way:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}.$$

The first subindex determines the row and the second subindex determines the column.

We will also write the matrix in short form by

$$A = \{a_{ij}\}_{1 \leq i \leq n, 1 \leq j \leq m}.$$

In this way we indicate that the subindex i varies from 1 to n and the subindex j varies from 1 to m. When we have fixed n, m we simply write $A = \{a_{ij}\}$.

A matrix of dimensions $1 \times n$, is called a *row vector*. A matrix of dimensions $n \times 1$, is called a *column vector*. For simplicity, we call just vector a column vector.

Example 1.2. The matrix $(3, 2, -1)$ is a row vector and the matrix $\begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix}$ is a column vector.

We denote by $M_{n \times m}(K)$ the set of all the matrices with coefficients in K having n rows and m columns.

By convention, we use capital letters to denote a matrix, and its entries are named with the same letter in lowercase.

Basic operations.

- *Identity matrix.* The identity matrix of order n is the matrix:

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

▶ *Sum.* Let $A, B \in M_{n\times m}(K)$, that is, let A, B be matrices of the same order $n \times m$. Then, one defines the sum of A and B, $C = A + B$ by:

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

That is, the term in the position (i, j) of the matrix $A + B$, is given by the sum of the entries in the position (i, j) of A and B:

$$C = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1m}+b_{1m} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2m}+b_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1}+b_{n1} & a_{n2}+b_{n2} & \cdots & a_{nm}+b_{nm} \end{pmatrix}.$$

For example, if $A = \begin{pmatrix} 3 & 2 & -1 \\ 1 & -4 & 1 \\ -6 & 2 & 3 \end{pmatrix}$, and $B = \begin{pmatrix} 0 & -1 & 1 \\ -2 & 3 & 4 \\ 1 & -2 & 2 \end{pmatrix}$. Then,

$$A + B = \begin{pmatrix} 3 & 1 & 0 \\ -1 & -1 & 5 \\ -5 & 0 & 5 \end{pmatrix}.$$

▶ *Transpose.* Let $A \in M_{n\times m}(K)$ be a matrix. Then, its transpose $A^t \in M_{m\times n}(K)$ is the matrix of order $m \times n$ with entry in the (i, j) position being a_{ji}. That is:

$$A^t = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{pmatrix}.$$

For example, the transpose of the matrix B above, is

$$B^t = B = \begin{pmatrix} 0 & -2 & 1 \\ -1 & 3 & -2 \\ 1 & 4 & 2 \end{pmatrix}$$

▶ *Product.* Let $A = \{a_{ij}\}$ be a matrix of order $n \times p$ and $B = \{b_{ij}\}$ be a matrix of order $p \times m$. Then, the product $C = A \cdot B = AB$ is the matrix with

$$c_{ij} = \sum_{k=1}^{p} a_{ip} b_{pj}.$$

That is, the element in the position (i, j) is obtained by the scalar product of the i-th row of the matrix A with the j-th column of B. **In order to be able to make the product of two matrices, it is necessary that the number of columns of the first matrix agrees with the number of rows of the second matrix.**

The product of the matrices A and B above is

$$A{\cdot}B = \begin{pmatrix} -5 & 5 & 9 \\ 9 & -15 & -13 \\ -1 & 6 & 8 \end{pmatrix}.$$

The *diagonal* of a matrix A consists of the elements a_{ij} with $i = j$. For instance, the diagonal of the matrix $\begin{pmatrix} 3 & 2 & -1 \\ 1 & -4 & 1 \\ -6 & 2 & 3 \end{pmatrix}$ is $(3, -4, 3)$.

Definition 1.3. Let $A \in M_{n\times m}(K)$ be a matrix of order $n \times m$. We say that A is a *square matrix* if $n = m$, that is, if it has the same number of rows and columns. In this case, we will just say that A is a *matrix of order* n. If A is square, then, we say:

- A is *symmetric*, if $A = A^t$. For instance, the matrix

$$\begin{pmatrix} 3 & 2 & -1 \\ 2 & -4 & -6 \\ -1 & -6 & 3 \end{pmatrix}$$

  is symmetric.

- A is *antisymmetric*, if $A = -A^t$. In particular, antisymmetric matrices have zeroes in the diagonal. For instance, the matrix

$$\begin{pmatrix} 0 & 2 & -1 \\ -2 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$$

  is antisymmetric.

1.2 Determinant of a matrix

We introduce here another operation, the determinant, that assigns to every **square matrix** A, a number $\det(A) \in K$. Among other uses, the determinant characterizes the matrices that are invertible.

Determinant of a matrix of order 2. Let A be a matrix of order 2×2, $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$. Then, the determinant of A is defined by

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Determinant of a matrix of order 3. Let A be a matrix of order 3×3, $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$. Then, the determinant of A is defined by

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{array}{l} a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ -a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}. \end{array}$$

Determinant of a matrix of order n. In general, let A be a square matrix of order n. Let A_{ij} be the matrix obtained by removing the ith row and the jth row of A. Fix any index j. Then, the determinant of A is defined recursively as follows:

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{vmatrix} = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \cdot \det(A_{ij}).$$

This is called to develop the determinant along the jth column. The result does not depend on the chosen column, hence, one should always take the one with more zeroes. Moreover, one could use a row instead of a column.

Properties of the determinant:

- If one matrix has two columns or two rows equal, then the determinant is zero.
- If the determinant of a matrix is zero, then it means that there is a linear combination of the columns equal to zero, and a linear combination of the rows equal to zero.
- If one column is a linear combination of the others, then the determinant is zero.
- The determinant of a matrix agrees with the determinant of its transpose.
- If there is one column or one row of zeroes in a matrix, then its determinant is zero.
- The determinant of a non-square matrix is not defined!

1.3 Rank and inverse

Definition 1.4. Let A be any matrix of order $n \times m$. A *minor of order* r of A is a square submatrix of A of order r. That is, a submatrix of A obtained by removing from A $n-r$ rows and $n-r$ columns.

The *rank* of A is r if any minor of order $> r$ has zero determinant and there exists a minor of order r whose determinant is different from zero. That is, the rank of A is the maximal order for which there exists a minor with determinant different from zero.

The rank of A is denoted by $\mathrm{rk}(A)$.

Example 1.5. Consider the matrix $A = \begin{pmatrix} 3 & 2 & -1 \\ 1 & -4 & 1 \\ 5 & -6 & 1 \end{pmatrix}$. If we compute the determinant, we obtain $\det(A) = 0$. Therefore, the rank of A is not three. To see if the rank is 2, we have to find a minor of order 2, whose determinant is different from zero. This is obtained by the minor $\begin{pmatrix} 3 & 2 \\ 1 & -4 \end{pmatrix}$. Hence, the rank of A is 2.

Inverse.

Definition 1.6. Let A be a square matrix of order n. We say that A is *invertible*, if there exists a matrix X such that

$$AX = I_n \quad \text{and} \quad XA = I_n.$$

In this case, we write $X = A^{-1}$.

Proposition 1.7. *A square matrix A is invertible if and only if* $\det(A) \neq 0$.

□

There are different methods to compute the inverse of a matrix A.

1.4 Systems of linear equations

Consider a system of linear equations $Ax = b$:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

The matrix A is called *the matrix of the system*, and the column vector b is called the *vector of independent terms*. The *augmented matrix* is the matrix obtained by joining the column vector b to the matrix A:

$$A|b = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2m} & b_2 \\ \vdots & \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} & b_n \end{pmatrix}$$

Not all systems have a unique solution. For instance, the system

$$\begin{aligned} 2x + 4y &= 0, \end{aligned}$$

has many solutions, namely, $(2, -1)$ is a solution, and all multiples of $(2, -1)$ are a solution. On the other hand, there are systems with no solution. For instance, they system

$$\begin{aligned} 2x + 4y &= 0, \\ x + 2y &= 3, \end{aligned}$$

has no solution.

In many occasions we need a criterium to decide whether a given system has or not a solution. One of the possible criterium is based on the rank of the matrix of the system A, and the rank of the augmented matrix $A|b$:

(i) If $\operatorname{rk}(A) = \operatorname{rk}(A|b)$, the system is called *compatible*, and there are solutions:

  (a) If the rank of A agrees with the number of variables (in this case, m), then there is a unique solution, and the system is called *determinate*.

  (b) Otherwise, the system is *indeterminate*.

(ii) If $\operatorname{rk}(A) \neq \operatorname{rk}(A|b)$, the system is called *incompatible*, and there are NO solutions.

Resolution. The standard methods to solve a system of linear equations "by hand", are the Cramer method and the Gauss method. The Gauss method consists of replacing the equations by a linear combination of the others, in such a way to obtain a linear system of equations, whose solutions are the same as the initial one, but with coefficient matrix being upper triangular.

2 Vector spaces

2.1 Vector spaces

A *vector space* over the real numbers $\mathbb{R}$, is a triple $(E, +, \cdot)$, where E is a set,

- $\triangleright$ $+$ is an internal operation (sum) of E, that is,

$$+ : E \times E \to E \qquad (u, v) \mapsto u + v.$$

- $\triangleright$ $\cdot$ is an external operation (scalar product) of E by $\mathbb{R}$, that is, there is a map

$$\cdot : \mathbb{R} \times E \to E \qquad (\lambda, u) \mapsto \lambda \cdot v.$$

The sum operation and the product by scalars must verify a list of properties (commutativity, associativity ...).

The elements of a vector space are called *vectors* and the elements of the $\mathbb{R}$ are called *scalars*.

The main example of a vector space is $\mathbb{R}^n$ with $n \in \mathbb{N}$. Recall that $\mathbb{R}^n$ is the set:

$$\mathbb{R}^n = \{(x_1, \ldots, x_n) |\ x_i \in \mathbb{R}\}.$$

Then, with the operations

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n),$$

and

$$\lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n), \qquad \lambda \in \mathbb{R},$$

$\mathbb{R}^n$ is a vector space over $\mathbb{R}$.

Other examples are the set of polynomials in one variable of degree less than or equal to a fixed d with coefficients in $\mathbb{R}$, $P_d[\mathbb{R}]$, and the set of matrices of order $n \times m$, $M_{n\times m}(\mathbb{R})$.

We will only consider vector spaces which are a subset of $\mathbb{R}^n$.

Definition 2.1. A *vector subspace* F of $\mathbb{R}^n$ is a non-empty set satisfying the following two conditions:

(i) $u + v \in F$ for all $u, v \in F$.

(ii) $\lambda v \in F$, for all $\lambda \in \mathbb{R}$ and $v \in F$.

Remark 2.2. If F is a subspace of $\mathbb{R}^n$, then F is itself a vector space over $\mathbb{R}$.

Example 2.3. Let $F = \{(x, y, z) \in \mathbb{R}^3 |\ x + z = 0\}$. Let us see that F is a vector subspace of E by checking the two conditions of last definition:

(i) Let $v_1 = (x_1, y_1, z_1), v_2 = (x_2, y_2, z_2) \in F$, i.e., $x_1 + z_1 = x_2 + z_2 = 0$. Then,

$$v_1 + v_2 = (x_1 + x_2, y_1 + y_2, z_1 + z_2).$$

To check that $v_1 + v_2 \in F$, we have to check that this element satisfies the condition defining F:

$$(x_1 + x_2) + (z_1 + z_2) = (x_1 + z_1) + (x_2 + z_2) = 0 + 0 = 0.$$

(ii) Let $v = (x, y, z) \in F$, i.e., $x + z = 0$. Then, given $\lambda \in \mathbb{R}$,

$$\lambda v = (\lambda x, \lambda y, \lambda z).$$

To check that $\lambda v \in F$, we have to check that this element satisfies the condition defining F:

$$\lambda x + \lambda z = \lambda(x + z) = \lambda \cdot 0 = 0.$$

Example 2.4. Consider $F = \{(x, y) \in \mathbb{R}^2 |\ xy = 0\}$. Then, F is NOT a vector subspace of $\mathbb{R}^2$ since:

$$(1, 0) \in F, \quad (0, 1) \in F,$$

but their sum is not in F

$$(1, 0) + (0, 1) = (1, 1) \notin F.$$

Remark 2.5. The vector subspaces of $\mathbb{R}^n$ are the ones defined by a collection of equations on $(x_1, \ldots, x_n)$, with no independent term, and no variables multiplied between them. This is what is called a *system of homogeneous linear equations*.

2.2 Linear independence, linear combination and basis

In this subsection we discuss the concepts of linear independence and linear combination of vectors and the concept of basis of $\mathbb{R}^n$.

Definition 2.6. Let $\mathcal{S} = \{v_1, \ldots, v_m\} \subset \mathbb{R}^n$ be a subset of vectors of $\mathbb{R}^n$.

(i) A vector $v \in \mathbb{R}^n$ *is a linear combination of* $v_1, \ldots, v_m$ if there exist scalars $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ such that

$$v = \lambda_1 v_1 + \cdots + \lambda_m v_m.$$

(ii) We denote by $\langle v_1, \ldots, v_m \rangle = \langle \mathcal{S} \rangle$ the set of all the linear combinations of $v_1, \ldots, v_m$. This is a vector subspace of $\mathbb{R}^n$.

(iii) If F is a subspace of $\mathbb{R}^n$ such that

$$F = \langle \mathcal{S} \rangle,$$

then we say that S *generates* F or that S *is a system of generators of* F. That is, $\{v_1, \ldots, v_m\}$ is a system of generators of F, if all vectors $u \in F$ can be written as a linear combination of $v_1, \ldots, v_m$.

Example 2.7. We have that $\mathbb{R}^2 = \langle (1, 0), (0, 1) \rangle$, since for all $(x, y) \in \mathbb{R}^2$, we have

$$(x, y) = x(1, 0) + y(0, 1).$$

But there are many other options. For example, we also have that $\mathbb{R}^2 = \langle (-1, 1), (1, 1) \rangle$, since for all $(x, y) \in \mathbb{R}^2$, we have

$$(x, y) = \frac{y - x}{2}(-1, 1) + \frac{x + y}{2}(1, 1).$$

Example 2.8. The set $\{(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)\}$ is a system of generators of $\mathbb{R}^n$.

Example 2.9. Let $F = \{(x, y, z) \in \mathbb{R}^3 |\ x + z = 0\}$ as in example 2.3. We know that it is a subspace of $\mathbb{R}^3$. We can find a system of generators as follows. From the equation $x + z = 0$, we have that $z = -x$. Hence, $(x, y, z) \in F$, if it is of the form $(x, y, -x)$, for some $x, y \in \mathbb{R}$. Then

$$(x, y, -x) = x(1, 0, -1) + y(0, 1, 0),$$

and we obtain that

$$F = \langle (1, 0, -1), (0, 1, 0) \rangle.$$

The last example gives a standard method to find a system of generators of a subspace given by a set of linear homogeneous equations. We just have to solve the system, leaving the solution depending on parameters.

Definition 2.10. Let $\mathcal{S} = \{v_1, \ldots, v_m\} \subset \mathbb{R}^n$ be a subset of vectors of $\mathbb{R}^n$. We say that $v_1, \ldots, v_m$ are *linearly independent* if the only solution of the linear combination

$$0 = \lambda_1 v_1 + \cdots + \lambda_m v_m,$$

is the trivial, that is, $\lambda_i = 0$ for all i. Otherwise, we say that the vectors are *linearly dependent*.

Example 2.11. The vectors $(1, 0), (0, 1) \in \mathbb{R}^2$ are linearly independent, since if we have

$$0 = \lambda_1(1, 0) + \lambda_2(0, 1) = (\lambda_1, \lambda_2),$$

it follows that $\lambda_1 = \lambda_2 = 0$.

The vectors $(-1, 1), (1, 1) \in \mathbb{R}^2$ are also linearly independent:

$$0 = \lambda_1(-1, 1) + \lambda_2(1, 1) = (-\lambda_1 + \lambda_2, \lambda_1 + \lambda_2)$$

implies that

$$-\lambda_1 + \lambda_2 = 0, \qquad \lambda_1 + \lambda_2 = 0,$$

and solving the system we see that $\lambda_1 = \lambda_2 = 0$.

Example 2.12. In $\mathbb{R}^n$, the vectors $(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$ are linearly independent.

Example 2.13. Continuing with example 2.3, check that the vectors $(1, 0, -1), (0, 1, 0) \in F$ are linearly independent.

Definition 2.14. We say that the set $\mathcal{S} = \{v_1, \ldots, v_m\}$ is a *basis* of F, if $v_1, \ldots, v_m$ are linearly independent and $\mathcal{S}$ is a system of generators of F.

Example 2.15. The set $\{(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)\}$ is a basis of $\mathbb{R}^n$. It is called the *canonical basis of* $\mathbb{R}^n$.

Example 2.16. Consider the vector space F of example 2.3. The vectors $(1, 0, -1), (0, 1, 0) \in F$ form a basis of F.

2.3 Dimension of a vector space

It is a fact that for every vector space, there exists a basis. Moreover, every basis has the same number of vectors.

Definition 2.17. Let F be a subspace of $\mathbb{R}^n$. The *dimension* of F is the number of elements of any of its basis.

Example 2.18. The dimension of $\mathbb{R}^n$ is n.

Example 2.19. Consider the vector space F of example 2.3. The dimension of F is 2.

Remark 2.20. There are vector spaces with infinite dimension, that is, vector spaces with no basis with finite number of vectors. This is the case, for instance, of the vector space of all continuous functions

$$f : \mathbb{R} \to \mathbb{R}.$$

We will not work with them in this course, but it is interesting to know that they exist.

The following is a very useful proposition, which easies the calculations many times.

Proposition 2.21. *Let F be a vector subspace of dimension m. Then,*

(i) Any set of linearly independent vectors of cardinal m is a basis of F.

(ii) Any system of generators of cardinal m is a basis of F.

□

Example 2.22. Consider the vector space F of example 2.3. The vectors $(1, 1, -1)$, $(1, -1, -1)$ belong to F. Since the dimension of F is 2, to see that they form a basis of F it is enough to check that they are linearly independent. That is, consider a linear combination

$$0 = \lambda_1(1, 1, -1) + \lambda_2(1, -1, -1).$$

This gives the system $\lambda_1 + \lambda_2 = 0;\ \lambda_1 - \lambda_2 = 0$ which leads to the solution $\lambda_1 = \lambda_2 = 0$. Therefore, these vectors are linearly independent and hence they form a basis of F.

2.4 Vector spaces and matrices

Once a basis of a vector space is fixed, we can reduce many operations and computations to matrix operations and to solve systems of linear equations.

Coordinates in a given basis.

Proposition 2.23. *Let $\mathcal{B} = \{v_1, \ldots, v_m\}$ be a basis of a vector subspace F. For any $u \in F$, there exist unique $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ such that*

$$u = \lambda_1 v_1 + \cdots + \lambda_m v_m.$$

We will write $u = (\lambda_1, \ldots, \lambda_m)_{\mathcal{B}}$ and call $(\lambda_1, \ldots, \lambda_m)$ the coordinates *of u in the basis $\mathcal{B}$.*

□

Example 2.24. The usual coordinates of a vector in $\mathbb{R}^n$, $u = (a_1, \ldots, a_n)$ are the coordinates of the vector u in the canonical basis.

Example 2.25. How do we express the vector $(1, 0, 0)$ in the basis

$$\mathcal{B} = \{(-1, 0, 1), (0, 1, 3), (2, 1, -1)\}?$$

We have to find x, y, z such that

$$(1, 0, 0) = x(-1, 0, 1) + y(0, 1, 3) + z(2, 1, -1).$$

This is equivalent to solve the system

$$\begin{aligned} 1 &= -x + 2z \\ 0 &= y + z \\ 0 &= x + 3y - z. \end{aligned}$$

The solution is $x = 0,\ y = \frac{1}{2},\ z = \frac{-1}{2}$. Hence, $(0, \frac{1}{2}, \frac{-1}{2})$ are the coordinates of $(1, 0, 0)$ in the basis $\mathcal{B}$.

Observe that the last example gives a method to find the coordinates of a vector u in a given basis $\mathcal{B}$. We are reduced to solve the system of linear equations with independent vector u and coefficient matrix the matrix whose columns are the vectors in the basis $\mathcal{B}$.

Checking if a collection of vectors are linearly independent. Now, imagine we are given some vectors $u_1, \ldots, u_r \in F$ and we want to see if they are linearly independent. Observe that necessarily, $r \leq n$. First of all, we write them in terms of the basis $\mathcal{B}$:

$$u_1 = (\lambda_{11}, \ldots, \lambda_{n1})_{\mathcal{B}}, \quad u_2 = (\lambda_{12}, \ldots, \lambda_{n2})_{\mathcal{B}}, \quad \ldots \quad u_r = (\lambda_{1r}, \ldots, \lambda_{nr})_{\mathcal{B}}.$$

In short, we would write

$$u_i = (\lambda_{1i}, \ldots, \lambda_{ni})_{\mathcal{B}},$$

for all i.

Then, consider the matrix with columns the coordinates of the vectors:

$$A = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1r} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2r} \\ \vdots & \vdots & & \vdots \\ \lambda_{n1} & \lambda_{n2} & \cdots & \lambda_{nr} \end{pmatrix}.$$

If the rank of A is r, then the vectors are linearly independent and they generate a vector subspace of dimension r. If the rank is less than r, let's say d, then the vectors are linearly dependent and they generate a vector subspace of dimension d.

Practical hints. Let us summarize the methods we have, in order to deal with vector spaces. During the class, we will try to understand these hints using several examples.

1. Given a set $\mathcal{S} = \{v_1, \ldots, v_m\}$ of vectors, how do we find a basis of the space $\langle \mathcal{S} \rangle$? Observe that, by definition, the vectors generate $\langle \mathcal{S} \rangle$, but they could be linearly dependent. We have to remove some vectors to obtain a system of linearly independent

vectors, but still generating $\langle \mathcal{S} \rangle$. But which vectors should we remove? First of all, compute the rank of the matrix whose columns are the coordinates of the vectors in $\mathcal{S}$. If the rank is r, then it means that we need to find only r linearly independent vectors in $\mathcal{S}$, to have a basis of $\langle \mathcal{S} \rangle$. For that, find a minor of order the rank of the matrix, whose determinant is non-zero. The column vectors involved in this minor are a basis of $\langle \mathcal{S} \rangle$.

2. If we are given a vector subspace F as the set of solutions of a homogeneous system of linear equations, how do we find a basis of F? First of all one has to solve the system, leaving the solution dependent on parameters. From that, we can find a general expression of an element in F and obtain a system of generators (see examples). From here we proceed as in the previous item.

3. Given a basis of a vector subspace $F = \langle v_1, \ldots, v_r \rangle$, how do I find a homogeneous system of linear equations defining F? We see it with an example. Let us take again our example 2.3:

$$F = \langle (1, 0, -1), (0, 1, 0) \rangle,$$

and let us forget that we already know one equation defining F. The vectors $(1, 0, -1)$, $(0, 1, 0)$ form a basis of F. A vector $(x, y, z) \in \mathbb{R}^3$ will belong to F, if it is a linear combination of $(1, 0, -1)$ and $(0, 1, 0)$. This is the same as to ask that the rank of the matrix

$$A = \begin{pmatrix} 1 & 0 & x \\ 0 & 1 & y \\ -1 & 0 & z \end{pmatrix}$$

is 2. So, we impose that the determinant of A is zero:

$$\det(A) = z + x = 0,$$

and we find the same equation that we had in the beginning.

This is a general method that works for arbitrary dimensions.

2.5 Change of basis

Consider two different basis of a vector subspace F of $\mathbb{R}^n$: $\mathcal{B}_e = \{e_1, \ldots, e_m\}$ and $\mathcal{B}_v = \{v_1, \ldots, v_m\}$.

Write every vector e_i as a linear combination of the vectors $v_1, \ldots, v_m$:

$$e_i = a_{1i} v_1 + \cdots + a_{mi} v_m.$$

Definition 2.26. The *matrix of base change from $\mathcal{B}_e$ to $\mathcal{B}_v$* is the matrix whose columns are the coordinates of the vectors e_i in the basis $\mathcal{B}_v$:

$$A_{\mathcal{B}_e \to \mathcal{B}_v} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{pmatrix}.$$

Example 2.27. Consider the two different basis of $\mathbb{R}^3$: $\mathcal{B}_v = \{(1,0,0),(0,1,0),(0,0,1)\}$ and $\mathcal{B}_e = \{(-1,0,1),(0,1,3),(2,1,-1)\}$ (exercise: check that they are basis of $\mathbb{R}^3$). Then, we have

$$A_{\mathcal{B}_e \to \mathcal{B}_v} = \begin{pmatrix} -1 & 0 & 2 \\ 0 & 1 & 1 \\ 1 & 3 & -1 \end{pmatrix}.$$

To find the matrix of base change from $\mathcal{B}_v$ to $\mathcal{B}_e$, we should find the coordinates of the vectors in $\mathcal{B}_v$ in the basis $\mathcal{B}_e$. Recall that this is found by solving a system of linear equations:

$$\begin{aligned} (1,0,0) &= x_1(-1,0,1) + y_1(0,1,3) + z_1(2,1,-1) \\ (0,1,0) &= x_2(-1,0,1) + y_2(0,1,3) + z_2(2,1,-1) \\ (0,0,1) &= x_3(-1,0,1) + y_3(0,1,3) + z_3(2,1,-1). \end{aligned}$$

Solving them, we find $x_1 = 0, y_1 = \frac{1}{2}, z_1 = -\frac{1}{2}, x_2 = -1, y_2 = \frac{1}{2}, z_2 = \frac{1}{2}, x_3 = \frac{1}{3}, y_3 = \frac{1}{6}, z_3 = -\frac{1}{6}$. Hence, we have

$$A_{\mathcal{B}_v \to \mathcal{B}_e} = \begin{pmatrix} 0 & -1 & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{6} \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{6} \end{pmatrix}.$$

Remark 2.28. Since the vectors $e_1, \ldots, e_m$ are linearly independent, the matrix $A_{\mathcal{B}_e \to \mathcal{B}_v}$ is a square matrix of maximum rank m, and hence its determinant is different from zero, and hence it is invertible.

Proposition 2.29. *The inverse of the matrix $A_{\mathcal{B}_e \to \mathcal{B}_v}$ is the matrix of base change from $\mathcal{B}_v$ to $\mathcal{B}_e$, denoted by $A_{\mathcal{B}_v \to \mathcal{B}_e}$. That is:*

$$A^{-1}_{\mathcal{B}_e \to \mathcal{B}_v} = A_{\mathcal{B}_v \to \mathcal{B}_e}.$$

Example 2.30. Check that the two matrices in the previous example are inverse to each other.

Why are these matrices useful? They allow to find the coordinates of a given vector in a new basis. That is, if

$$v = x_1 e_1 + \cdots + x_m e_m = (x_1, \ldots, x_m)_{\mathcal{B}},$$

the coordinates of v in the basis $v_1, \ldots, v_m$ are obtained by the matrix product

$$A_{\mathcal{B}_e \to \mathcal{B}_v} \cdot v = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}.$$

Example 2.31. Consider the basis

$$\mathcal{B}_v = \{(1,0,0),(0,1,0),(0,0,1)\}$$

and

$$\mathcal{B}_e = \{(-1,0,1),(0,1,3),(2,1,-1)\}$$

of $\mathbb{R}^3$. The coordinates of the vector $(2,0,-1)$ in the basis $\mathcal{B}_v$ are found as follows:

$$\begin{pmatrix} 0 & -1 & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{6} \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{6} \end{pmatrix} \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{3} \\ \frac{5}{6} \\ -\frac{7}{6} \end{pmatrix}.$$

3 Diagonalization

Definition 3.1. Two matrices $A, B \in M_{n\times n}(\mathbb{R})$ are *equivalent* if there exists an invertible matrix $C \in M_{n\times n}(\mathbb{R})$, such that:

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

Proposition 3.2. *If we have an equality $A = CBC^{-1}$, then*

(i) $\det(A) = \det(B)$.

(ii) $\mathrm{rk}(A) = \mathrm{rk}(B)$.

Definition 3.3. A matrix A is *diagonalizable* if it is equivalent to a diagonal matrix D, i.e., if there exists a diagonal matrix D and an invertible matrix C such that

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

Diagonal matrices are interesting for many reasons. One of them is because it's very simple to make computations with them. For instance, if D is diagonal,

$$D = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{pmatrix},$$

then, the m-power of D is just the m-power of the elements in the diagonal

$$D^m = \begin{pmatrix} d_1^m & 0 & \dots & 0 \\ 0 & d_2^m & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^m \end{pmatrix}.$$

For a general matrix, to compute A^m is hard if m is very big. But if A is diagonalizable, then we have $A = CDC^{-1}$, with D diagonal, and then,

$$A^m = (CDC^{-1})^m = CDC^{-1}CDC^{-1}\cdot\ldots\cdot CDC^{-1}CDC^{-1} = CD^mC^{-1},$$

which as we saw it is easy to compute.

Not only this, there are many situations when to know that a matrix is diagonalizable and to know the matrix C and the elements $d_1, \dots, d_n$ of the diagonal matrix, allows one to obtain some nice properties. We will see some examples during the course. For the moment, we focus on understanding when a matrix is diagonalizable and how to compute the matrices C and D.

When is a matrix diagonalizable? Consider a square matrix A of order n and the matrix $A - xId$, obtained by removing the indeterminate x from all entries in the diagonal. For example, if

$$A = \begin{pmatrix} 2 & 0 & 4 \\ 3 & -4 & 12 \\ 1 & -2 & 5 \end{pmatrix},$$

then

$$A - xId = \begin{pmatrix} 2-x & 0 & 4 \\ 3 & -4-x & 12 \\ 1 & -2 & 5-x \end{pmatrix}.$$

The determinant of the matrix $A - xId$ is a polynomial of degree n, called the *characteristic polynomial of* A, and denoted by $c_A(x)$.

In our example,

$$\det(A - xId) = \begin{vmatrix} 2-x & 0 & 4 \\ 3 & -4-x & 12 \\ 1 & -2 & 5-x \end{vmatrix} = -x(x^2 + 3x - 2).$$

Next, we look for the zeroes of this polynomial $c_A(x)$, that is, the values of x for which $c_A(x) = 0$. Be aware that the zeroes can be complex numbers!

Definition 3.4. The zeroes of the characteristic polynomial $c_A(x)$ are called the *eigenvalues* of A.

In our example, the solutions for $-x(x^2 + 3x - 2) = 0$ are $x = 0, 1, 2$. Therefore, the eigenvalues of A are $0, 1, 2$.

Let us think about what we are doing. The eigenvalues, are the values of x for which the determinant of $A - xId$ is zero. This is the same as to say that they are the values for which the rank of the matrix $A - xId$ is less than n.

Consider the homogeneous system of linear equations

$$(A - xId)v = 0.$$

It will have solutions different from zero only if the system is indeterminate, that is, only when the rank of the coefficients matrix $A - xId$ is less than n. Therefore, the eigenvalues are the values of x for which there exist vectors $v \neq 0$ such that $(A - xId)v = 0$.

If we rewrite this equality we see that $(A - xId)v = 0$ if and only if

$$Av = xv.$$

Therefore, λ is an eigenvalue if there exists $v \in \mathbb{R}^n$ such that

$$Av = \lambda v.$$

In this case, the vector v is called an *eigenvector*.

The eigenvectors of an eigenvalue λ are found by solving the system

$$(A - \lambda Id)v = 0.$$

The set of the solutions form a vector subspace of $\mathbb{R}^n$, denoted by E_λ. The dimension of this space is always equal to or greater than 1, and equal to or smaller than the multiplicity of λ as a zero of the polynomial $c_A(x)$.

If we denote by $\mu_A(\lambda)$ the multiplicity of λ as a zero of $c_A(x)$, we are saying that

$$1 \leq \dim E_\lambda \leq \mu_A(\lambda),$$

for every eigenvalue λ.

Proposition 3.5. *A matrix real A of order n is diagonalizable if, and only if, all the zeros of $c_A(x)$ are real, and*

$$\dim E_\lambda = \mu_A(\lambda),$$

for all eigenvalues λ.

Proposition 3.6. *If A is diagonalizable, then the set formed by a basis of each space of eigenvectors E_λ is a basis of $\mathbb{R}^n$.*

Therefore, the process to decide whether a matrix is diagonalizable consists of the following:

(i) Compute the characteristic polynomial of A.

(ii) Find the zeroes of the characteristic polynomial, that is, the eigenvalues.

(iii) If there are complex eigenvalues, the matrix is not diagonalizable in $\mathbb{R}$. If this is not the case, solve the system of equations

$$(A - \lambda Id)v = 0,$$

for all eigenvalues λ, and find a basis of the solution.

(iv) If the dimensions of E_λ coincide with the multiplicities of λ, then the matrix diagonalizes. If not, the matrix is not diagonalizable.

A particular case is the case when the characteristic polynomial $c_A(x)$ decomposes in linear factors, all different. That is, all the zeroes are real, and the multiplicity of all of them is 1. Then, by the inequalities $1 \leq \dim E_\lambda \leq \mu_A(\lambda) = 1$, we see, without computing $\dim E_\lambda$, that necessarily $\dim E_\lambda = 1 = \mu_A(\lambda)$, and by our criteria, the matrix diagonalizes.

Proposition 3.7. *If all zeroes of the characteristic polynomial of A are real and all of them have multiplicity one, then the matrix is diagonalizable.*

If the matrix is diagonalizable, then the diagonal matrix D is obtained by putting the eigenvalues along the diagonal. The matrix C is obtained by putting the eigenvectors in columns. One has to keep the same order for the eigenvalues and for the eigenvectors. In fact, C is the matrix of basis change from the basis formed by the eigenvectors to the canonical basis of $\mathbb{R}^n$.

Remark 3.8. If A is diagonalizable, then the determinant of A is the product of the eigenvalues of A, each of them to the power given by its multiplicity in the characteristic polynomial.

Remark 3.9. Assume that A is a square matrix of order n. Then, the rank of A is r if and only if 0 is an eigenvalue of multiplicity $n - r$.

Examples. Let us see how this works by checking several examples.

Example 3.10. Let us start with the example we had from above. In that case, we had that the eigenvalues were $\lambda = 0, 1, 2$. Since they are all real and have multiplicities 1, the matrix is diagonalizable. The matrix D to which A diagonalizes, is

$$D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

To find the matrix C, we need to solve the appropriate systems of equations:

- The space of eigenvectors for the eigenvalue 0, E_0, is obtained by solving the system $Av = 0$, with $v = (x, y, z)$:

$$\begin{pmatrix} 2 & 0 & 4 \\ 3 & -4 & 12 \\ 1 & -2 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0,$$

  that is, the system

$$2x + 4z = 0, \qquad 3x - 4y + 12z = 0, \qquad x - 2y + 5z = 0.$$

  We have that $x = -2z$ and hence from the third equation $2y = x + 5z = 3z$. Hence,

$$E_0 = \{(x, y, z) |\ x = -2z, y = \frac{3}{2}z\} = \{(-2z, \frac{3}{2}z, z)\} = \langle(-2, \frac{3}{2}, 1)\rangle = \langle(-4, 3, 2)\rangle.$$

- The space of eigenvectors for the eigenvalue 1, E_1, is obtained by solving the system $(A - Id)v = 0$, with $v = (x, y, z)$:

$$\begin{pmatrix} 1 & 0 & 4 \\ 3 & -5 & 12 \\ 1 & -2 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0,$$

  that is, the system

$$x + 4z = 0, \qquad 3x - 5y + 12z = 0, \qquad x - 2y + 4z = 0.$$

  We have that $x = -4z$ and hence from the third equation $2y = x + 4z = 0$. Hence,

$$E_1 = \{(x, y, z) |\ x = -4z, y = 0\} = \{(-4z, 0, z)\} = \langle(-4, 0, 1)\rangle.$$

- The space of eigenvectors for the eigenvalue 2, E_2, is obtained by solving the system $(A - 2Id)v = 0$, with $v = (x, y, z)$:

$$\begin{pmatrix} 0 & 0 & 4 \\ 3 & -6 & 12 \\ 1 & -2 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0,$$

  that is, the system

$$z = 0, \qquad 3x - 6y + 12z = 0, \qquad x - 2y + 3z = 0.$$

  We have that $x = 2y$. Hence,

$$E_2 = \{(x, y, z) |\ z = 0, x = 2y\} = \{(2y, y, 0)\} = \langle(2, 1, 0)\rangle.$$

Hence, the matrix C is

$$C = \begin{pmatrix} -4 & -4 & 2 \\ 3 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix}.$$

Check that it is satisfied $A = CDC^{-1}$.

Example 3.11. Consider now the matrix $A = \begin{pmatrix} -4 & 0 & 2 \\ 0 & 1 & 0 \\ 5 & 1 & 3 \end{pmatrix}$. The characteristic polynomial is

$$c_A(x) = \det(A - xId) = \begin{vmatrix} -4-x & 0 & 2 \\ 0 & 1-x & 0 \\ 5 & 1 & 3-x \end{vmatrix} = -x^3 + 3x - 2 = -(x-1)^2(x+2).$$

The zeroes of this polynomial are: $\lambda = -2, 1$, the eigenvalue -2 with multiplicity 1 and the eigenvalue 1 with multiplicity 2. A is diagonalizable only if the dimension of E_1 is 2. Let us check it: The space of eigenvectors for the eigenvalue 1, E_1, is obtained by solving the system $(A - Id)v = 0$, with $v = (x, y, z)$:

$$\begin{pmatrix} -5 & 0 & -2 \\ 0 & 0 & 0 \\ 5 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0,$$

that is, the system

$$-5x - 2z = 0, \qquad 5x + y + 2z = 0.$$

We have that $y = 0$ and $z = -\frac{5}{2}x$. Hence,

$$E_1 = \{(x, y, z) |\ y = 0, z = -\frac{5}{2}x\} = \{(x, 0, z) = -\frac{5}{2}x)\} = \langle (2, 0, -5) \rangle.$$

The dimension of E_1 is 1, and hence A is not diagonalizable.

Example 3.12. Consider the matrix $A = \begin{pmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{pmatrix}$. The eigenvalues (exercise) are $\lambda = -1$ with multiplicity 2 and $\lambda = 8$ with multiplicity 1. Let us check the dimension of E_{-1}. We solve the system

$$\begin{pmatrix} 4 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0,$$

that is, the system $2x + y + 2z = 0$. We have that $y = -2x - 2z$ and hence

$$E_{-1} = \{(x, y, z) |\ y = -2x - 2z\} = \{(x, -2x - 2z, z)\} = \langle (1, -2, 0), (0, -2, 1) \rangle.$$

Since the dimension is 2, the matrix A is diagonalizable.

With octave. Let A be a square matrix of order n. The command

$$[V, x] = eig(A),$$

stores in V the eigenvectors of the matrix A and in x the eigenvalues of the matrix A. A is diagonalizable if the number of vectors in V is n.

4 Linear maps

Let F, G be vector spaces. A map $f : F \to G$ is called a *linear map*, if:

(i) For all $u, v \in F$, we have $f(u+v) = f(u) + f(v)$.

(ii) For all $\lambda \in \mathbb{R}$ and $v \in F$, we have $f(\lambda v) = \lambda f(v)$.

Given a matrix A, of order $n \times m$, the product of the matrix with a vector defines a linear map

$$\begin{array}{ccc} \mathbb{R}^m & \xrightarrow{f_A} & \mathbb{R}^n \\ v & \mapsto & Av. \end{array}$$

Written in coordinates, if $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}$, and $v = (x_1, \ldots, x_m) \in \mathbb{R}^m$, we have

$$Av = (a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m, \ldots, a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m) \in \mathbb{R}^n.$$

For example, consider the matrix

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

Then, it defines the linear map

$$\begin{array}{ccc} \mathbb{R}^3 & \xrightarrow{f_A} & \mathbb{R}^3 \\ (x, y, z) & \mapsto & (-4x - 4y + 2z, 3x + z, 2x + y). \end{array}$$

In practice, all linear maps are the product by a matrix. They are always described by homogeneous linear equations on $x_1, \ldots, x_n$.

Diagonalization and linear maps. Consider a matrix A and the associated linear map f_A. Then, the eigenvectors are the vectors that are mapped to a multiple of itself by f_A.

5 Orthogonal basis and symmetric matrices

5.1 Orthogonal basis

Orthogonality. Let $v_1 = (x_1, \ldots, x_n)$, $v_2 = (y_1, \ldots, y_n)$ be vectors of $\mathbb{R}^n$. Consider the *scalar product* of vectors in $\mathbb{R}^n$, given by:

$$< v_1, v_2 >= v_1 \cdot v_2 = (x_1, \ldots, x_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = x_1 y_1 + \cdots + x_n y_n = \sum_{i=1}^{n} x_i y_i.$$

We denote this product with a dot "$\cdot$" or brackets $<, >$.

For example, if $v_1 = (1, -1, 2)$ and $v_2 = (0, 2, -1)$, the scalar product of v_1 and v_2 is:

$$v_1 \cdot v_2 = (-1)2 + 2(-1) = -4.$$

Definition 5.1. We say that two vectors v_1, v_2 are *orthogonal*, if

$$v_1 \cdot v_2 = 0.$$

For example, the vectors $v_1 = (1, -1, 2)$ and $v_2 = (0, 2, 1)$ are orthogonal, since

$$v_1 \cdot v_2 = (-1)2 + 2 \cdot 1 = -2 + 2 = 0.$$

Definition 5.2. A basis $\{e_1, \ldots, e_n\}$ of a vector space F is called a *orthogonal basis* if all vectors are orthogonal to each other, that is

$$e_i e_j = 0 \qquad i \neq j.$$

Example 5.3. The canonical basis of $F = \mathbb{R}^n$ is an orthogonal basis.

The basis of $\mathbb{R}^2$ given by $\{(1, -1), (1, 1)\}$ is orthogonal.

Coordinates in an orthogonal basis. Fix an orthogonal basis $\mathcal{B} = \{e_1, \ldots, e_n\}$ of a vector space F. Then, it is very easy to find the coordinates of a vector $v = (x_1, \ldots, x_n) \in F$ in this basis.

Recall that to find the coordinates of v in the basis $\mathcal{B}$ means to find the scalars $a_1, \ldots, a_n$ for which

$$v = a_1 e_1 + \cdots + a_n e_n. \tag{5.4}$$

If the basis is orthogonal, these are given by the formula

$$a_i = \frac{v \cdot e_i}{e_i \cdot e_i}.$$

Indeed, if we consider the scalar product by e_i at both sides of equation (5.4), we obtain:

$$v \cdot e_i = a_1 e_1 \cdot e_i + \cdots + a_i e_i \cdot e_i + \cdots + a_n e_n \cdot e_i.$$

Since the basis is orthogonal, $e_i \cdot e_j = 0$ if $i \neq j$, and hence we have

$$v \cdot e_i = a_i e_i \cdot e_i,$$

and we obtain the formula given above.

Remark 5.5. Orthogonality is the algebraic concept corresponding to perpendicularity. Hence, a basis is orthogonal if the vectors are perpendicular to each other. The coordinates of a vector v in a orthogonal basis correspond to the orthogonal projection of the vector to each of the lines given by each of the vectors.

Orthonormality. The square root of the scalar product of a vector with itself is called its *norm*. That is:

$$||x|| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

Definition 5.6. A basis $\mathcal{B} = \{e_1, \ldots, e_n\}$ of a vector space F is called orthonormal if

$$e_i \cdot e_j = \begin{cases} 0 & i \neq j, \\ 1 & i = j. \end{cases}$$

That is, if the vectors are orthogonal to each other and moreover each vector has norm 1.

Example 5.7. The canonical basis of $\mathbb{R}^n$ is orthonormal. The basis $\{(1,-1),(1,1)\}$ of $\mathbb{R}^2$ is orthogonal but not orthonormal, since

$$||(1,-,1)|| = \sqrt{1+1} = \sqrt{2} \neq 1.$$

To obtain an orthonormal basis, we just divide each vector by its norm. Hence, the basis

$$\{(1/\sqrt{2}, -1/\sqrt{2}), (1/\sqrt{2}, 1/\sqrt{2})\}$$

is orthonormal.

If $\{e_1, \ldots, e_n\}$ is a orthonormal basis of a vector space F and v is any vector, then the coordinates of v in the basis $\{e_1, \ldots, e_n\}$ are given by

$$a_i = v \cdot e_i.$$

Remark 5.8. A orthonormal basis of $\mathbb{R}^n$ is always obtained by rotating the canonical basis of $\mathbb{R}^n$, along different axis of rotation through the origin $(0, \ldots, 0)$.

Gramm-Schmidt. Given an orthogonal basis of F, $\{e_1, \ldots, e_n\}$, we can always obtain a orthonormal basis by diving each vector by its norm. That is, the basis

$$\left\{ \frac{e_1}{||e_1||}, \ldots, \frac{e_n}{||e_n||} \right\}$$

is orthonormal.

More in general, given an arbitrary basis $\mathcal{B} = \{v_1, \ldots, v_n\}$, there is a process, called the *Gram-Schmidt orthonormalization*, to obtain an orthonormal basis of F from $\mathcal{B}$.

The importance of this fact is that, first of all, there exist orthonormal basis for any vector subspace F, and second, not only we know that they exist, but we know how to compute one if we have any basis.

This is performed with Octave with the command:

Vout = GramSchmidt(V)

Orthogonal projections. Let e be a vector in $\mathbb{R}^n$. Then, the multiples of e define a line through the origin. Let v be any vector. Then, the *orthogonal projection of* v to the line with direction e is the vector

$$\mathrm{Proj}_e(v) = \frac{v \cdot e}{e \cdot e}\, e.$$

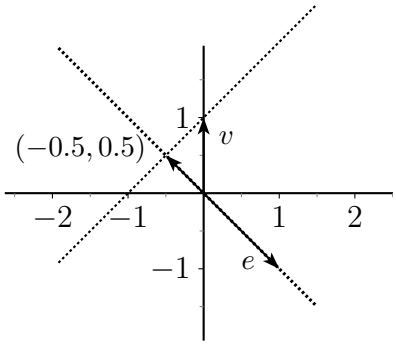
Example 5.9. Consider the vector $e = (1, 0, 1)$ and $v = (0, 1, 1)$ of $\mathbb{R}^3$. The orthogonal projection of v into e is

$$\text{Proj}_e(v) = \frac{(0,1,1)\cdot(1,0,1)}{(1,0,1)\cdot(1,0,1)}\ (1,0,1) = \frac{1}{2}(1,0,1) = (1/2, 0, 1/2).$$

Example 5.10. Consider the vector $e = (1, -1)$ and $v = (0, 1)$ of $\mathbb{R}^2$. The orthogonal projection of v into the line with direction e is

$$\text{Proj}_e(v) = \frac{(0,1)\cdot(1,-1)}{(1,-1)\cdot(1,-1)}\ (1,-1) = \frac{-1}{2}(1,-1) = (-1/2, 1/2).$$

Graphically, we have



More in general, consider a subspace $F \subset \mathbb{R}^n$ and a vector $v \in \mathbb{R}^n$, not in F. How do we compute the orthogonal projection of v into F?

The easiest way is to consider an orthogonal basis of F, $\{v_1, \ldots, v_r\}$. Then,

$$\text{Proj}_F(v) = \frac{v\cdot v_1}{v_1\cdot v_1}\ v_1 + \cdots + \frac{v\cdot v_r}{v_r\cdot v_r}\ v_r.$$

Orthogonal matrices. Let A be an invertible square matrix. Then, we say that A is *orthogonal*, if

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

That is, the inverse of A is exactly its transpose.

Proposition 5.11. *The basis change matrix between two orthonormal basis is always orthogonal.*

As a consequence of the last proposition, we see that if $\{e_1, \ldots, e_n\}$ is a orthonormal basis of $\mathbb{R}^n$, then the matrix whose columns are the coordinates of this vector in the canonical basis, is orthogonal. Indeed, this matrix is the matrix of basis change between $\{e_1, \ldots, e_n\}$ to the canonical basis of $\mathbb{R}^n$.

Hence, the matrix

$$\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

is orthogonal. Actually, this is the matrix corresponding to a rotation of angle 45^o.

5.2 Symmetric matrices

Symmetric matrices. A special case of matrices are the real symmetric matrices. They are interesting in the sense that they are ALWAYS diagonalizable, and moreover, there exists an orthonormal basis of eigenvectors.

Recall that a symmetric matrix satisfies that it is equal to its transpose $A = A^t$.

Proposition 5.12. *Let A be a symmetric matrix of order n. Then, A is diagonalizable. Moreover, eigenvectors of different eigenvalue are orthogonal, that is, if v is an eigenvector for λ and w is an eigenvector for μ, with $\mu \neq \lambda$, then*

$$v \cdot w = 0.$$

Let A be a symmetric matrix and consider an orthonormal basis for each space of eigenvectors E_λ. Since eigenvectors of different eigenvalue are orthogonal, the union of these basis gives an orthonormal basis of $\mathbb{R}^n$.

By proposition 5.11, the matrix C whose columns are given by the eigenvectors is orthogonal, and hence $C^{-1} = C^t$. Hence, the equivalency of matrices is given by

$$A = CDC^t,$$

with D a diagonal matrix.

Bilinear forms. Consider a symmetric matrix A of order n and the following map:

$$\begin{array}{rcl} \mathbb{R}^n \times \mathbb{R}^n & \xrightarrow{g_A} & \mathbb{R} \\ (v, w) & \mapsto & v^t A w. \end{array}$$

The maps of this form are called *symmetric bilinear*, and they satisfy the following properties:

(i) $g_A(u+v, w) = g_A(u, w) + g_A(v, w)$, $g_A(u, v+w) = g_A(u, v) + g_A(u, w)$ for all $u, v, w \in \mathbb{R}^n$.

(ii) $g_A(\lambda v, w) = \lambda g_A(v, w)$, $g_A(v, \mu w) = \mu g_A(v, w)$, for all $v, w \in \mathbb{R}^n$ and $\lambda, \mu \in \mathbb{R}$.

(iii) $g_A(v, w) = g_A(w, v)$ (commutativity).

Any map satisfying the previous properties is given by the product by a symmetric matrix as above.

Observe that the maps g_A can be considered as a sort of scalar product, since they assign to every pair of vectors a scalar in $\mathbb{R}$. The matrix associated to the usual scalar product is the identity.

Definition 5.13. We will say that v, w are orthogonal with respect to A, if $g_A(v, w) = 0$. The norm with respect to A of a vector v will be the square root of $g_A(v, v)$.

Example 5.14. For example, consider the matrix

$$A = \begin{pmatrix} 1 & -1 \\ -1 & 0 \end{pmatrix}.$$

The map g_A is the one that assigns to every pair of vectors $(x_1, y_1), (x_2, y_2)$, the scalar:

$$(x_1, y_1) \begin{pmatrix} 1 & -1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = (x_1, y_1) \begin{pmatrix} x_2 - y_2 \\ -x_2 \end{pmatrix} = x_1 x_2 - x_1 y_2 - x_2 y_1.$$

We know that all real symmetric matrices diagonalize. Hence, let A be a symmetric matrix and let v be an eigenvector of eigenvalue λ_v and w be an eigenvector of eigenvalue λ_w and assume that $\lambda_v \neq \lambda_w$. Then,

$$g_A(v, w) = v^t A w = v^t(\lambda_w) w = \lambda_w v \cdot w = 0,$$

since eigenvectors of different eigenvalue are orthogonal, and hence, by definition, their scalar product is zero.

Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis of eigenvectors for A. We have

$$g_A(v_i, v_i) = v_i^t A v_i = v_i^t(\lambda_{v_i}) v_i = \lambda_{v_i} v_i \cdot v_i = \lambda_{v_i},$$

since $v_i \cdot v_i = 1$. Hence, the orthonormal basis $\{v_1, \ldots, v_n\}$ is also orthogonal with respect to A and the norms with respect to A of the eigenvectors are the square roots of the eigenvalues of A.

To sum up:

- All real symmetric matrices A are diagonalizable. We can choose the basis in which they diagonalize to be orthonormal. Then, we have an equality of matrices

  $$A = UDU^t,$$

  with U an orthogonal matrix and D a diagonal matrix.

- The basis of eigenvectors obtained $\{u_1, \ldots, u_n\}$, satisfies the following property:

  $$u_i^t A u_j = \begin{cases} 0 & i \neq j, \\ \lambda_i & i = j, \end{cases}$$

  if λ_i is the eigenvalue corresponding to the eigenvector u_i.

Singular Value Decomposition. Let X be a real matrix, non-necessary square, of order $n \times m$, with $n \geq m$, that is, with more rows than columns. Then, there exists always a decomposition of A of the form

$$X = UDV^t,$$

where

- D is a diagonal matrix of order $n \times n$,
- U is a matrix of order $n \times m$, whose columns are orthogonal to each other and have norm 1. This is satisfied if $U^t U = id$.
- V is a orthogonal matrix of order $m \times m$.

This is called the *singular value decomposition* of X.

From the previous results, we can see how it is obtained. To illustrate the explanations, we will follow an example. So, consider the matrix:

$$X = \begin{pmatrix} 5 & 0 \\ 0 & 5 \\ -4 & 3 \end{pmatrix}.$$

Consider the matrix obtained by making the product of X^t by X, $A = X^tX$. Since the order of X^t is $m \times n$ and the order of X is $n \times m$, the order of A is $m \times m$. Observe that

$$A^t = (X^tX)^t = X^t(X^t)^t = X^tX = A.$$

Therefore, the matrix A is symmetric.

In our example, we have

$$A = X^tX = \begin{pmatrix} 5 & 0 & -4 \\ 0 & 5 & 3 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & 5 \\ -4 & 3 \end{pmatrix} = \begin{pmatrix} 41 & -12 \\ -12 & 34 \end{pmatrix}.$$

Since the matrix is symmetric, it diagonalizes in an orthonormal basis $\{v_1, \ldots, v_m\}$. Let V be the matrix whose columns are the vectors $\{v_1, \ldots, v_m\}$. We know that this is a orthogonal matrix. Let D be the diagonal matrix of the eigenvalues. Hence, we have

$$A = VDV^t.$$

In our example, we find that the eigenvalues are

$$D = \begin{pmatrix} 25 & 0 \\ 0 & 50 \end{pmatrix}.$$

The eigenvectors are given by

$$E_{25} = \langle (3,4) \rangle, \qquad E_{25} = \langle (-4,3) \rangle.$$

As it should be, the vectors $(3,4), (-4,3)$ are orthogonal, but they are not of norm 1. Therefore, to obtain the matrix V, we should normalize them dividing by their norm. In this way, we obtain the orthonormal basis of $\mathbb{R}^2$:

$$\{(3/5, 4/5), (-4/5, 3/5)\}.$$

The matrix V is then given by

$$V = \begin{pmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{pmatrix}.$$

Then, consider the matrix

$$U = XVD^{-1/2}.$$

This is a matrix of order $n \times m$. The columns of U are orthogonal to each other and has norm 1. Indeed:

$$\begin{aligned} U^tU &= (XUD^{-1/2})^tXUD^{-1/2} = D^{-1/2}U^tX^tXUD^{-1/2} \\ &= D^{-1/2}U^tUDU^tUD^{-1/2} = D^{-1/2}DD^{-1/2} = id. \end{aligned}$$

From the relation $U = XVD^{-1/2}$, it follows that $X = UD^{1/2}V$ as desired.

In our example,

$$U = \begin{pmatrix} 5 & 0 \\ 0 & 5 \\ -4 & 3 \end{pmatrix} \begin{pmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & \sqrt{50} \end{pmatrix} = \begin{pmatrix} 3/5 & -2\sqrt{2}/5 \\ 4/5 & 3\sqrt{2}/10 \\ 0 & \sqrt{2}/2 \end{pmatrix}$$

Analogously, consider the product $B = XX^t$. From the description above, we have

$$XX^t = UD^{1/2}V^t(UD^{1/2}V^t)^t = UD^{1/2}V^tVD^{1/2}U^t = UDU^t.$$

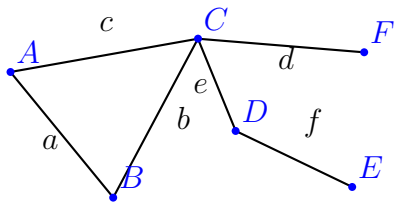
We see that D is also the diagonal matrix associated to the symmetric matrix XX^t.

6 Graphs and measures of centrality

6.1 First definitions on graphs

Definition 6.1. A *graph* G is a collection of entities with relations between them.
The entities of a graph are called *nodes or vertices*.

The set of nodes is denoted by $V(G)$ and the set of edges is denoted by $E(G)$.
Graphically, a graph is represented by drawing a dot for each node, and a line between two dots if they are related. For example:



This is a graph with 6 vertices $V(G) = \{A, B, C, D, E, F\}$ and 6 edges $E(G) = \{a, b, c, d, e, f\}$.

Definition 6.2. The *degree* of a node is the number of nodes to which it is connected by an edge. For a graph G and a vertex v, we denote it by $d_G(v)$.

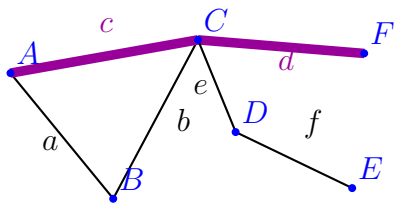
In our example,

$$d_G(A) = d_G(B) = d_G(D) = 2, \quad d_G(C) = 4, \quad d_G(E) = d_G(F) = 1.$$

Definition 6.3. Consider a graph G and two nodes v_1, v_2.

- A *path* between v_1 and v_2 is a collection of edges $e_1, \ldots, e_r$ starting at v_1 and ending at v_2.
- The *length* of a path is the number of edges in it.
- A path is called *geodesic* if it has minimal length, that is, all other paths between the two nodes have at least the same length.

For example, the darkened edges form a path of length 2 between A and F, the path cd:



There is still another path between A and F, abd, which has length 3. There is only one geodesic path between A and F, which is cd.

Definition 6.4. Consider a graph G and two nodes v_1, v_2. The *distance* between the nodes v_1 and v_2 is the minimal length of a path connecting them. It is denoted by $\delta_G(v_1, v_2)$ or simply $\delta(v_1, v_2)$.

In our example, the distance between A and F is 2. The shortest path between them is cd. That is, $\delta(A, F) = 2$.

We have

$$\begin{aligned}\delta(A,C) = \delta(A,B) = \delta(B,C) = \delta(C,D), \delta(C,F) = \delta(D,E) = 1,\\ \delta(A,F) = \delta(B,F) = \delta(A,D) = \delta(B,D) = \delta(C,E) = \delta(D,F) = 2,\\ \delta(A,E) = \delta(B,E) = \delta(F,E) = 3.\end{aligned}$$

6.2 Measures of centrality

We give here the classical measures of centrality of a node in a graph. Each of them follows a different principle of what it means to be central.

Degree centrality. The degree centrality is based on the following assumption:

The most important node in a graph is the one connected to more nodes.

So, the *degree centrality* of a node v is given by its degree:

$$C_d(v) = d_G(v).$$

Then, the more central nodes are the ones with higher degree centrality. In our example, the more central node, according to the degree centrality is C.

Closeness centrality. This measure of centrality is based in the following assumption:

An important node is typically close to, and can communicate quickly with, the other nodes in the network.

By this assumption, there are a few different definitions of closeness centrality that could be made. We just give the one called *eccentricity centrality*.

Given a graph G and a node v, the *eccentricity centrality* of v is the higher distance between v and the rest of the nodes:

$$C_e(v) = \max_{w \in V(G)} \delta(v, w).$$

In this case, the more central nodes will be the ones with minimal eccentricity centrality.

In our example, we have

$$C_e(A) = C_e(B) = C_e(E) = C_e(F) = 3, \qquad C_e(C) = C_e(D) = 2.$$

By this criterium, the more central nodes are C and D. Observe that with the degree centrality, D was not consider top central.

Betweenness centrality. For this measure, the assumption made is the following:

An important node will lie on a high proportion of paths between other nodes in the network.

That is, there will be many paths going through this node, so that if you remove it, many paths are broken.

Consider a graph G and a node v. For two different nodes u, w, different from v, define σ_{uw} to be the number of geodesic paths between u and w, and $\sigma_{uw}(v)$ be the number of geodesic paths between u and w that pass through v. Then, the *betweenness centrality* is defined by:

$$C_b(v) = \sum_{u \neq w} \frac{\sigma_{uw}(v)}{\sigma_{uw}}.$$

The higher the betweenness centrality, the more central is the node.

In our example, the betweenness centrality of C is the higher and its value is 8. For the node D it is 4, while for the nodes A, B, E, F is 0. So, the more central node here is C.

Eigenvector centrality. Here, the assumption being made is the following:

An important node is connected to important neighbors.

Let us index the nodes in the form $v_1, \ldots, v_m$, if there are m nodes in the graph. In our example, we fix $v_1 = A, v_2 = B, \ldots$ and so on.

For this measure, we want to assign a score x_i to each node v_i, that measures the assumption made. For that, we want the score x_i of the node v_i to be related to the score of its neighbors, in a way that if the neighbors have high score, then this node has also high score. This is formalized by saying that the score of a node should be proportional to the sum of the scores of its neighbors:

$$x_i = K \sum_{v_j \text{ neighbor of } v_i} x_j, \tag{6.5}$$

for all i.

Up to now, we haven't defined the values x_i, we are just writing down what they should satisfy.

In order to understand better how to solve the previous relation, we introduce the *adjacency matrix* A of a graph G. The entry (i, j) of this matrix is 1 if there in an edge between the nodes v_i, v_j and it is 0 otherwise. The diagonal entries are zeroes.

It has, therefore, as many rows and columns as the number of nodes. Moreover, it is always symmetric.

In our example, the adjacency matrix is:

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

Let $x = (x_1, \ldots, x_m)$ be the vector of the scores. With the adjacency matrix, the equation (6.5) is rewritten as

$$x = KAx,$$

or in other words

$$Ax = \lambda x,$$

for $\lambda = \frac{1}{K}$. So, the scores we are looking for are given by an eigenvector of the adjacency matrix. If we impose some conditions, like having all scores positive, we determine the eigenvector. Observe that the adjacency matrix is diagonalizable because it is symmetric.

The condition of all scores being positives is not always possible to solve, but there are many theorems related to their existence (this is called the *Perron-Frobenius theory*).

What we can assure, though, is that the wanted eigenvector corresponds to the higher positive eigenvalue of the adjacency matrix.

If we diagonalize the adjacency matrix of our example, we find that the eigenvalues are $-1.92, -1, -0.75, 0.29, 1, 2.3$. The highest eigenvalue is 2.3. The space of eigenvectors corresponding to the eigenvalue 2.3 is

$$E_{2.3} = \langle (0.46, 0.46, 0.63, 0.32, 0.13, 0.26) \rangle.$$

We can take any multiple of this vector as the score system.

In this case, this means that

$$x_1 = x_2 = 0.46,\ x_3 = 0.63,\ x_4 = 0.32,\ x_5 = 0.13,\ x_6 = 0.26.$$

The highest score is for the third node, that is C. Observe that the nodes A and B have the same score. This makes a lot of sense due to the symmetry in the graph of this two nodes.

7 Introduction to ordinary differential equations

7.1 Introduction

Population growth. Consider the problem of studying the way in which the size of a population varies. For that, let $p(t)$ be the number of individuals in a population at time t.

At time $t+T$, the number of individuals will be $p(t+T)$, so the number of added individuals is $p(t+T)-p(t)$. Assuming that the bigger is T, the more individuals are expected to arrive, and the shorter the time the fewer individuals are expected to arrive, we could write the change $p(t+T)-p(t)$ by

$$p(t+T)-p(t)=NT, \text{ which is } \frac{p(t+T)-p(t)}{T}=N.$$

Now, taking the limit as $T\to 0$, we obtain that the function $p(t)$ satisfies the equality:

$$\frac{dp}{dt}=\lim_{T\to 0}\frac{p(t+T)-p(t)}{T}=N.$$

To be more realistic, assume that the factor N is a function of the time t, and the size of the population at time t, so that we have

$$\frac{dp(t)}{dt}=N(t,p(t)).$$

What we got is an equation relating a function $p(t)$ with its derivative, that is, with its variation. The functions $p(t)$ satisfying the equation describe the growth of the population by telling us at every time t which is the size of the population.

We start by considering the simplest case: $N(t,p(t))=N_0p(t)$. That is, we assume that the increase of population at time t is linear with respect to the size of the population at time t. Then, we have the equation

$$\frac{dp}{dt}=N_0p(t).$$

To find a solution, we rewrite the equation in the form

$$\frac{dp}{p(t)}=N_0dt.$$

Integrating both sides of the equality from time 0 to time t, we obtain

$$\int_0^t\frac{dp(t)}{p(t)}=N_0\int_0^t dt.$$

Observe that

$$\int_0^t\frac{dp(t)}{p(t)}=\int_{p(0)}^{p(t)}\frac{dp}{p}.$$

Solving both sides, we obtain:

$$\log(p(t))\big|_0^t=N_0t\big|_0^t$$

which gives

$$N_0t=\log\left(\frac{p(t)}{p(0)}\right),$$

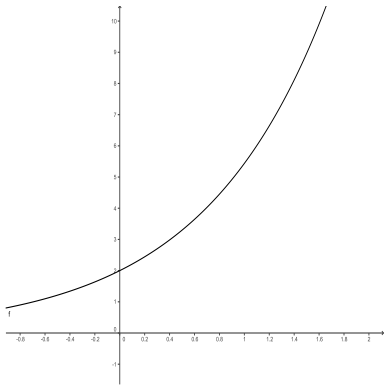


Figure 1: Exponential growth: $N_0 = 1$ and $p(0) = 2$.

and so

$$p(t) = p(0)e^{N_0 t}.$$

So, the solution curve is an exponential depending on the initial size of the population $p(0)$. The curve is plotted in figure 7.1, for $N_0 = 1$ and $p(0) = 2$.

This is obviously a non-realistic description of the growth of a population. Nevertheless, it describes quite well the early stages of the growth of a population.

A more realistic point of view should take at least into account the deaths and even the migration events. Another model of the growth of a population is given by the Verhulst's equation:

$$\frac{dp}{dt} = N_0 p(t) - D_0 p(t) + I(t) - E(t),$$

where N_0 is the birth rate, D_0 the death rate, $I(t)$ is the number of immigrants at time t while $E(t)$ is the number of emigrants at time t.

But... how to solve this equation? Which are the functions $p(t)$ satisfying such an equation? It is not easy anymore.

Drug administration. When a drug is administered, it forms a concentration in the body fluids. This concentration diminishes in time through elimination, destruction or inactivation.

The rate of reduction of the concentration is usually found to be proportional to the concentration. We therefore get the following relation:

$$\frac{dc(t)}{dt} = -\frac{c(t)}{\tau},$$

where $c(t)$ denotes the concentration at time t and τ is a constant.

The solution of this equation is

$$c(t) = c_0 e^{-t/\tau},$$

with c_0 the concentration at time t. We see that the larger τ is, the slower the drug disperses.

Imagine now that the drug is administered at different times $t = 0, t_0, 2t_0, \ldots$ with the same dose c_0.

At time t_0, just before the second dose is administered, the leftover drug in the body will be

$$r_1 = c_0 e^{-t_0/\tau}.$$

This concentration is added to the dose C_0 for the second dose:

$$c_1 = c_0 + c_0 e^{-t_0/\tau}.$$

The drug left just before the third dose at time $2t_0$ is

$$r_2 = c_1 e^{-t_0/\tau} = c_0 e^{-t_0/\tau} + c_0 e^{-t_0/\tau} e^{-t_0/\tau} = c_0 e^{-t_0/\tau}(1 + e^{-t_0/\tau}),$$

so after the third dose is administered, the concentration c_2 is

$$c_2 = c_0 + r_2 = c_0(1 + e^{-t_0/\tau} + e^{-2t_0/\tau}).$$

In general, at time $t = (n-1)t_0$, we have

$$c_{n-1} = c_0(1 + e^{-t_0/\tau} + e^{-2t_0/\tau} + \cdots + e^{-(n-1)t_0/\tau}) = c_0 \frac{1 - e^{-nt_0/\tau}}{1 - e^{-t_0/\tau}}.$$

Observe that when n tends to infinity, the term $e^{-nt_0/\tau}$ tends to zero, and hence

$$c_M = \lim_{n\to\infty} c_n = \frac{c_0}{1 - e^{-t_0/\tau}},$$

which tells us that the amount of drug in the body will never exceed this upper bound.

These studies are useful to determine the frequency of the doses and the concentration c_0 for a treatment.

Cell division. We can apply the considerations on the growth of a population to the cell division process. In this case, we should take into account the crowding effects, that could induce lack of nutrient, shortage of oxygen or other effects.

To model cell division, one takes into account that if we have p cells, there are p^2 possible interactions. In this way, one can write:

$$\frac{dp}{dt} = N_0 p(t) - a p(t)^2,$$

with N_0, a positive constants. The term $N_0 p(t)$ accounts for the increase of cells due to division. The term $a p(t)^2$ accounts for the inhibition growth due to crowding effects.

The equation

$$\frac{dp}{dt} = N_0 p(t) - a p(t)^2,$$

is called the *differential equation of logistics*.

The solution of this equation is given by

$$p(t) = \frac{N_0 p(0)}{a p(0) + (N_0 - a p(0)) e^{-N_0 t}}.$$

This is known as the *logistic law of growth*.

Newton's law. Newton's second law tells us that the acceleration of a body at time t equals the product of the mass of the body by the force that is affecting it. The acceleration of a body is given by the second derivative of the position $x(t)$. So, we obtain the equation

$$\frac{d^2x}{dt^2} = mF(x(t)).$$

If we are considering the movement of a pendulum, we obtain the differential equation

$$\frac{d^2x}{dt^2} = -\sin(x)$$

(normalized so that the mass of the body is 1).

The movement of a spring, is given by the differential equation

$$\frac{d^2x}{dt^2} = -x.$$

7.2 Ordinary differential equations

All the examples above give an unknown function as a solution of an equation involving the derivatives of the functions. These equations are called differential equations. In this section we give the formalism for the study of differential equations.

Definition 7.1. An *ordinary differential equation (ODE)* is an equation of the form

$$F\left(t, y, \frac{dy}{dt}, \frac{d^2y}{dt^2}, \ldots, \frac{d^ny}{dt^n}\right) = 0,$$

with $y = y(t)$ a function of t. That is, an equation relating a function depending on one variable t and its higher derivatives.

Remark 7.2. One also writes $y^{(r)}$ instead of $\frac{d^ry}{dt^r}$. The variable t is called the *independent variable* and y is called the *dependent variable*.

It is also common to write $\dot{y}$ instead of y', for the first derivative.

Example 7.3. The equations given in the introductory section are all ordinary differential equations.

Definition 7.4. The *order* of an ordinary differential equation is the order of the highest derivative appearing in the equation.

For instance, the order of the ODE

$$\frac{d^4y}{dt^4} = \frac{dy}{dt}\left(\frac{d^3y}{dt^3}\right)^4 + y^2$$

is 4, and the order of the ODE

$$\left(\frac{d^2y}{dt^2}\right)^3 = t^2 + y^2$$

is 2.

Definition 7.5. An ODE is *autonomous* if it is not explicitly dependent on time t, that is, if it is of the form

$$F\left(y, \frac{dy}{dt}, \frac{d^2y}{dt^2}, \ldots, \frac{d^ny}{dt^n}\right) = 0.$$

For example, the ODE

$$y' = y^2 + t$$

is not autonomous, while the differential equation

$$y' = y^2$$

is autonomous.

General and particular solution. A solution is a function $y : \mathbb{R} \to \mathbb{R}$ satisfying the given ODE.

Given an ODE, it might have or not a solution. Under some favorable conditions, we can ensure that it has a solution.

The *general solution* of an ODE of order n is the solution of the ODE, containing n constants. It gives the general form of any solution of the ODE. A particular solution is a solution of the ODE.

Particular solutions are found by fixing some *initial conditions*.

For instance, consider the ODE

$$\frac{dy}{dt} = y.$$

The general solution is $y(t) = Ce^t$, with C a constant. A particular solution is $y(t) = e^t$, and it is obtained by knowing that $y(0) = 1$.

In general, it is hard to solve an ODE, even knowing that a solution exists. There are several analytic methods to solve ODE's, but in many occasions one is forced to use numerical methods to find approximations of the solution.

Nevertheless, sometimes even when one is able to find a solution, the expression of the solution is extremely difficult to interpret. The qualitative study of differential equations aims to study the main features of the solution function, without solving the differential equation.

Linear ordinary differential equations. A *linear ordinary differential equation* is an ODE of the form

$$a_n(t)\frac{d^ny}{dt^n} + a_{n-1}(t)\frac{d^{n-1}y}{dt^{n-1}} + \cdots + a_1(t)\frac{dy}{dt} + a_0(t)y = f(t), \tag{7.6}$$

with $a_0(t), \ldots, a_n(t)$ known functions of t. If these functions are constant, then we say that the equation is a *linear ordinary differential equation with constant coefficients*.

If an ordinary differential equation is not of the form (7.6), we call it a non-linear ODE.

For example,

$$y' + y'' = 0$$

is a linear differential equation with constant coefficients. The differential equation

$$y'' + 2ty' = 0$$

is a linear differential equation, but with non-constant coefficients. The differential equation

$$(y')^2 + y'' = 0$$

is non-linear, because there is a non-linear term: $(y')^2$.

Linearization of an autonomous ODE of order 1. Consider the ODE

$$y' = f(y),$$

and consider the first order Taylor approximation of the function $f(y)$ at a fixed point a:

$$f(y) \sim f(a) + f'(a)(y - a).$$

The linear approximation of the ODE $y' = f(y)$ at $y = a$ is then the linear ODE

$$y' = f(a) + f'(a)(y - a).$$

For example, consider the ODE

$$y' = y^2 - y.$$

We have $f(y) = y^2 - y$ and hence $f'(y) = 2y - 1$. The linear approximation at $a = 0$ is

$$y' = -y,$$

and at $a = 1$ is

$$y' = y - 1.$$

Phase space. Consider an autonomous differential equation of order 1, $y' = f(y)$. The *phase space* of a differential equation is the space of solution values $y(t)$ of the differential equation.

If we have a differential equation of order 1, for example

$$\frac{dy}{dt} = y - y^3 = f(y),$$

the phase space is a one-dimensional line with coordinate y:

That is, we think of $y(t)$ as the position of a particle moving along the axis at some time t. The differential equation tells us that the derivative at the position $y(t)$ with respect to t is exactly $f(y(t))$. If $f(y(t)) > 0$, then it means that $y(t)$ is increases, while if $f(y(t)) < 0$, we have that $y(t)$ decreases. If $f(y(t))$ is zero, then the derivative of $y(t)$ is zero and hence the particle does not move.

The points such that $y' = 0$, are called *equilibrium points*. If y_1 is such a point, then the constant function

$$y(t) = y_1$$

is a solution of the differential equation.

Definition 7.7. ▶ An equilibrium point is called *stable*, if the solutions near the point tend to the point as t tends to infinity.

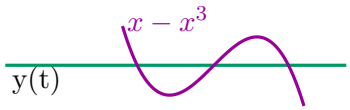
▶ An equilibrium point is called *unstable*, if the solutions near the point scape from the point as t tends to infinity.

▶ Otherwise, the point is called *critic*.

How do we detect the stability? Consider our example

$$\frac{dy}{dt} = y - y^3 = f(y),$$

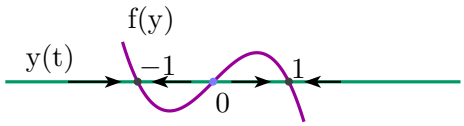
and we plot the phase space and the function $f(y)$:



The intersection of the curve $f(y)$ with y, gives the points where $y'(t) = 0$, and hence the equilibrium points. In our example, these are

$$y = 0, 1, -1.$$

Observe that in the interval $(-\infty, -1) \cup (0, 1)$, the value of $f(y)$ is positive. Hence, a particle in that position moving along a solution of the ODE, increases the position. In the interval $(-1, 0) \cup (1, +\infty)$, the function $f(y)$ is negative, and hence a particle in this interval decreases the position. This is shown graphically as:



We see that the solutions near $y = -1, 1$, approach the point. These points are therefore stable equilibrium points. The point $y = 0$ is an unstable equilibrium point.

The general rule to determine the stability of an equilibrium point is as follows.

Proposition 7.8. *Consider a differential equation of order 1 of the form*

$$y' = f(y),$$

and let y_0 be an equilibrium point.

(i) If $f'(y_0) < 0$, then the point is stable.

(ii) If $f'(y_0) > 0$, the the point is unstable.

Let us consider again the differential equation of logistics and try to understand better the concept of phase space. Recall that the equation is given by

$$\frac{dp}{dt} = N_0 p(t) - ap(t)^2,$$

and the solution of this equation is

$$p(t) = \frac{N_0 p(0)}{ap(0) + (N_0 - ap(0))e^{-N_0 t}}.$$

To simplify the notation, let us fix $a = p(0) = 1$, that is, we start with one cell, and let $N_0 = 2$.

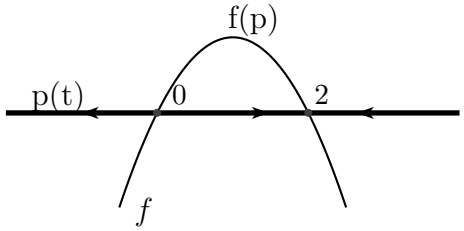
So, we have that the differential equation is

$$\frac{dp}{dt} = 2p(t) - p(t)^2,$$

and the solution is

$$p(t) = \frac{2}{1 + e^{-2t}}.$$

The phase space of the differential equation is:



We see that $0, 2$ are equilibrium points. The point 0 is obvious, since if we start with no cell, we stay with no cell. This point is unstable. The equilibrium point 2 is stable, since all solutions tend to 2.

Let us do it with the constants a and N_0 and assume that $N_0 > a$. Then, the equilibrium points of the ODE are given by

$$0 = N_0 p - ap^2 = p(N_0 - ap).$$

Hence, the equilibrium points are $p = 0$ and $p = N_0/a$. The sign of the differential tells us that N_0/a is a stable equilibrium point. Observe the importance of this fact: no matter what is the initial population of cells, this model tells us that as time passes, we tend to have N_0/a cells.

The plot of the solution for the constants we fixed previously is shown in figure 2.

We see that the function approaches asymptotically to $y = 2$. This is the same conclusion we had reached without solving the ODE!

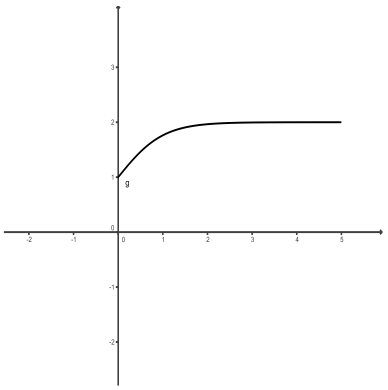


Figure 2: Exponential growth: $N_0 = 2$, $a = 1$ and $p(0) = 1$.

7.3 Dimension 2 systems of ODE's

Volterra-Lotka model. We saw how the growth of a population can be described in terms of an ordinary differential equation. Imagine the more complex situation where one species is the "prey" of another species (the "predator"). Then, the predator species has an inhibiting effect on the growth of the prey and the prey has an accelerating effect on the growth of the predator.

We denote by $x(t)$ the size of the prey population and by $y(t)$ the size of the predator population at time t. Let us assume that without the existence of the predator, the prey population would increase following a logistic growth curve, that is

$$\frac{dx}{dt} = ax - bx^2, \qquad a, b > 0.$$

Assuming that the predation effect is proportional to the prey population and to the predator population, we could write

$$\frac{dx}{dt} = ax - bx^2 - cxy, \qquad a, b, c > 0.$$

For the predator species, in the absence of prey, we can assume that the population would decrease, due to the lack of nutrients. Then, we have

$$\frac{dy}{dt} = -ey + fxy, \qquad e, f > 0.$$

So, we have that the population sizes $x(t)$ and $y(t)$ depend on each other, and the variation is described by a 2-dimensional system of ordinary differential equations:

$$\begin{aligned} \frac{dx}{dt} &= ax - bx^2 - cxy, \\ \frac{dy}{dt} &= -ey + fxy. \end{aligned}$$

A simpler model can be obtained by assuming that the size of the prey population follows an exponential growth instead of the logistic growth:

$$\begin{aligned} \frac{dx}{dt} &= ax - cxy, \\ \frac{dy}{dt} &= -ey + fxy. \end{aligned}$$

Epidemics. The spread of an epidemics can also be studied by a system of differential equations. The simplest way to model it is the following.

Let $S(t)$ be the number of individuals susceptible to catch the illness, and let $M(t)$ be the number of ill individuals. We could assume the following:

- The rate at which the susceptible individuals catch the illness is proportional to the number of ill people and to the number of susceptible people.
- The rate of ill people that die or get cured is proportional to the number of ill people.

Then, the spread of the epidemics can be described by the differential equations

$$\begin{aligned} \frac{dS}{dt} &= -kMS, \\ \frac{dM}{dt} &= kMS - aM. \end{aligned}$$

Some definitions. A *system of ordinary differential equations* of order n is a system of equations of the form:

$$\begin{aligned} y_1^{(n)} &= f_1(t, y, y', \ldots, y^{(n-1)}), \\ \vdots &\quad \vdots \\ y_m^{(n)} &= f_m(t, y, y', \ldots, y^{(n-1)}), \end{aligned}$$

for $y(t) = (y_1(t), \ldots, y_m(t)) : \mathbb{R}^m \to \mathbb{R}$.

The system is called *linear* if all the differential equations are linear ODE's.

We will restrict to 2-dimensional systems of order 1 and *autonomous*:

$$\begin{aligned} x' &= f(x, y), \\ y' &= g(x, y), \end{aligned}$$

for functions $x(t), y(t) : \mathbb{R} \to \mathbb{R}$ and $f, g : \mathbb{R}^2 \to \mathbb{R}$.

Example 7.9. The following are autonomous systems of ordinary differential equations of order 1:

$$\begin{cases} x' &= x^2 - y, \\ y' &= \log(xy), \end{cases} \qquad \text{and} \qquad \begin{cases} y' &= x - y, \\ x' &= x + y. \end{cases}$$

Linearization of a 2-dimensional system of ODE's of order 1. The linearization of an autonomous ordinary differential equation of order 1 explained above is extended to the systems of ordinary differential equations, by considering the Taylor approximation of the functions $f(x, y)$ and $g(x, y)$ at a given point (a, b).

Therefore, consider the system of ODE's

$$\begin{aligned} x' &= f(x, y), \\ y' &= g(x, y). \end{aligned}$$

The first order Taylor approximation of f at (a, b) is

$$f(x, y) \sim \frac{\partial f}{\partial x}\Big|_{(a,b)}(x - a) + \frac{\partial f}{\partial y}\Big|_{(a,b)}(y - b)$$

and the first order Taylor approximation of g at (a,b) is

$$g(x,y) \sim \frac{\partial g}{\partial x}\Big|_{(a,b)}(x-a) + \frac{\partial g}{\partial y}\Big|_{(a,b)}(y-b).$$

Then, the linear system of ODE's associated to our system is the system:

$$\begin{aligned} x' &= \frac{\partial f}{\partial x}\Big|_{(a,b)}(x-a) + \frac{\partial f}{\partial y}\Big|_{(a,b)}(y-b), \\ y' &= \frac{\partial g}{\partial x}\Big|_{(a,b)}(x-a) + \frac{\partial g}{\partial y}\Big|_{(a,b)}(y-b). \end{aligned}$$

Observe, that written in matrix notation, this system is

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} f(a,b) \\ g(a,b) \end{pmatrix} + \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix}_{|(a,b)} \begin{pmatrix} x-a \\ y-b \end{pmatrix}.$$

The coefficient matrix is called the *Jacobian matrix* at (a,b). The general form of the Jacobian matrix is the matrix

$$\begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix}$$

For instance, consider the system

$$\begin{aligned} x' &= xy + 2x, \\ y' &= y^3 + x^2, \end{aligned}$$

and $a = (0,1)$. Then, the linear approximation of the system is the system

$$\begin{aligned} x' &= 3x, \\ y' &= 3(y-1). \end{aligned}$$

The Jacobian matrix is in this case

$$\begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}.$$

Remark 7.10. In general, for a system of ODE's of the form:

$$\begin{aligned} y_1' &= f_1(y_1, \ldots, y_m), \\ \vdots &\quad \vdots \\ y_m' &= f_m(y_1, \ldots, y_m), \end{aligned}$$

the linearization at a point $a = (a_1, \ldots, a_m)$ is the linear system of ODE's

$$\begin{pmatrix} y_1' \\ y_2' \\ \vdots \\ y_m' \end{pmatrix} = \begin{pmatrix} f_1(a) \\ f_2(a) \\ \vdots \\ f_m(a) \end{pmatrix} + \begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \frac{\partial f_1}{\partial y_2} & \cdots & \frac{\partial f_1}{\partial y_m} \\ \frac{\partial f_2}{\partial y_1} & \frac{\partial f_2}{\partial y_2} & \cdots & \frac{\partial f_2}{\partial y_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1} & \frac{\partial f_m}{\partial y_2} & \cdots & \frac{\partial f_m}{\partial y_m} \end{pmatrix}_{\Big|(a_1,\ldots,a_m)} \begin{pmatrix} y_1 - a_1 \\ y_2 - a_2 \\ \vdots \\ y_m - a_m \end{pmatrix}.$$

The coefficient matrix is called the *Jacobian matrix* of the ODE at $a = (a_1, \ldots, a_n)$.

7.4 Stability of a 2-dimensional system

In this section we give an introduction to the stability analysis of 2-dimensional systems of autonomous differential equations. That is, we will try to get a picture of the solutions of the system, without knowing them.

Phase space and trajectory. Consider a 2-dimensional autonomous system of ordinary differential equations

$$\begin{aligned} x' &= f(x,y), \\ y' &= g(x,y). \end{aligned}$$

Let $(h_1(t), h_2(t))$ be a pair of solutions of the system. Then, we can plot the points $x = h_1(t)$ and $y = h_2(t)$ in the (x,y)-plane at time t. As t varies, the points (x,y) will trace a curve in the plane, which is known as a *trajectory*. The (x,y)-plane is called the *phase plane* or *phase space*.

Observe that for a given solution, the vector (x', y') gives the vector tangent to the trajectory at every point (x,y). The slope of this vector is then

$$\frac{y'}{x'} = \frac{dy}{dx} = \frac{g(x,y)}{f(x,y)}.$$

So, the trajectories are the curves whose tangent vector, at every point (x,y), is given by $(f(x,y), g(x,y))$.

We see that if we can solve the ordinary differential equation

$$\frac{dy}{dx} = \frac{g(x,y)}{f(x,y)},$$

(called *the implicit equation*) we can trace the trajectories in the phase space.

For instance, consider the system

$$\begin{aligned} x' &= y, \\ y' &= -2x. \end{aligned}$$

The implicit equation is

$$\frac{dy}{dx} = \frac{-2x}{y},$$

that is

$$ydy = -2xdx.$$

Integrating both sides from $x = x_0$ to x, we get

$$\int_{y_0}^{y} ydy = \int_{x_0}^{x} -2xdx.$$

This is:

$$\frac{y^2}{2}\Big|_{y_0}^{y} = -x^2\Big|_{x_0}^{x}, \qquad \frac{y^2}{2} + x^2 = C,$$

with $C = x_0^2 + \frac{y_0^2}{2}$. We get a different curve for any different initial conditions (x_0, y_0). Observe that all of them are ellipses. A picture of the phase space is drawn below. Can you tell the direction of the trajectories?

Under favorable conditions (for the existence of solutions), two different trajectories do not intersect.

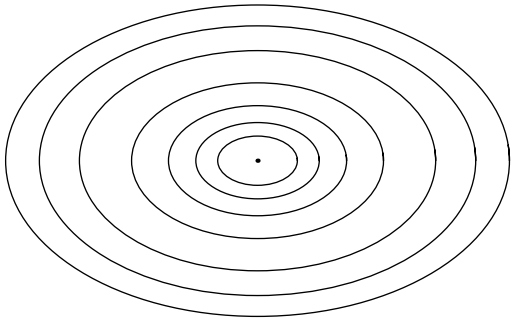


Figure 3: Phase space of the system $x' = y$, $y' = -2x$.

Equilibrium points and local stability.

Definition 7.11. An *equilibrium point or steady state* is a solution (x, y) of a system of ODE's such that

$$f(x, y) = 0, \qquad g(x, y) = 0.$$

An equilibrium point gives a constant trajectory which is a solution of the system.

As in the 1-dimensional case, we can distinguish the equilibrium points in three groups, the stable, the unstable and the critical. An equilibrium point is (locally) *stable* if all solutions near the point tend to the point when t tends to infinity. If the solutions near the point move away from the point, then the point is called (locally) *unstable*. Otherwise it is called critical.

How can we check if a point is stable or unstable? Here is the criterium.

Proposition 7.12. *Consider a 2-dimensional system of differential equations*

$$\begin{aligned} x' &= f(x, y), \\ y' &= g(x, y). \end{aligned}$$

Let (a, b) be an equilibrium point and let $J_{(a,b)}$ be the Jacobian of the system at the point (a, b). Then:

(i) If all the eigenvalues of $J_{(a,b)}$ have negative real part, then the point (a, b) is stable.

(ii) If there is at least one eigenvalue of $J_{(a,b)}$ with positive real part, then the point (a, b) is unstable.

(iii) Otherwise, the point (a, b) is critical.

Example 7.13. Consider the system of differential equations

$$\begin{aligned} x' &= y - x^2 + x, \\ y' &= x - y. \end{aligned}$$

The equilibrium points are obtained by solving the system

$$\begin{aligned} 0 &= y - x^2 + x, \\ 0 &= x - y. \end{aligned}$$

We see that $x = y$ and then from the first equation,

$$-x^2 + 2x = 0$$

we have that the solutions are $(0, 0)$ and $(2, 2)$.

To determine the stability of each of the points, we need to consider the Jacobian at each of the points.

The general form of the Jacobian is

$$J = \begin{pmatrix} -2x + 1 & 1 \\ 1 & -1 \end{pmatrix}$$

For $(a, b) = (0, 0)$, we obtain

$$J_{(0,0)} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

The eigenvalues of this matrix are $-\sqrt{2}, \sqrt{2}$. Since one of them has positive real part, the equilibrium point $(0, 0)$ is unstable.

For $(a, b) = (2, 2)$, we have

$$J_{(2,2)} = \begin{pmatrix} -3 & 1 \\ 1 & -1 \end{pmatrix}.$$

The eigenvalues of this matrix are $-3.41..., -0.58...$ Since both of them are negative, the equilibrium point is stable.

Null clines.

Definition 7.14. The curves obtained by setting

$$f(x, y) = 0, \qquad \text{or} \qquad g(x, y) = 0$$

are called the *x-nullcline and y-nullcline* respectively.

Observe that the equilibrium points are obtained by intersecting the two nullclines. The study of the sign of $f(x, y), g(x, y)$ in the regions delimited by the nullclines give a general picture of the global stability of the solution trajectories.

Example 7.15. The nullclines of the system

$$\begin{aligned} x' &= y - x^2 + x, \\ y' &= x - y, \end{aligned}$$

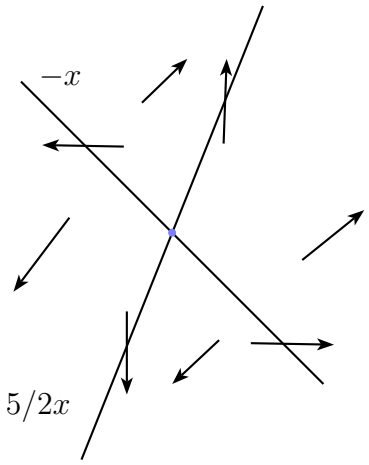
are the line $x = y$ and the parabola $y = x^2 - x$.

How can we get a general picture of the solution trajectories of a system of differential equations? Remember that what we know is the value of the tangent line at each point of the trajectory. To simplify, we can think that the direction of the tangent vector is described by the sign of the two coordinates, that is, knowing at what quadrant it points. For a change of sign, we need to cross a point where one of the two coordinates is zero, that is we need to cross a nullcline. Therefore, we can have a general picture of the solution of a system of differential equations by diving the phase space in regions delimited by the nullclines and checking the sign of the functions $f(x, y)$ and $g(x, y)$ at any point in the region.

Example 7.16. Consider now the system

$$\begin{aligned} x' &= 5x + 2y, \\ y' &= 2x + 2y. \end{aligned}$$

The only equilibrium point is $(0, 0)$, which is unstable, the nullclines are $x = -y$ and $x = \frac{2}{5}y$. The general picture of the phase space is given by:



where we have checked the signs of the tangent vectors considering:

$p = (1, 0)$: $v = (5, 2)$.

$p = (0, 1)$: $v = (2, 2)$.

$p = (-1, 0)$: $v = (-5, -2)$.

$p = (0, -1)$: $v = (-2, -2)$.

For the example 7.15, the obtained picture is shown in figure 7.4. We can see that the point $(2, 2)$ is globally stable.

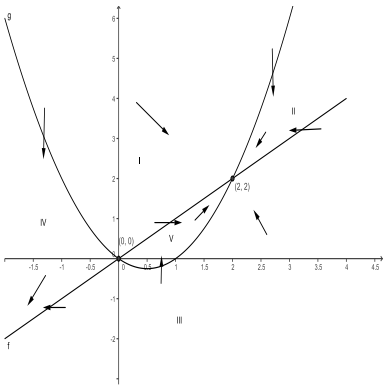


Figure 4: Phase space of the system in example 7.15.

Oscillations. A solution $(x(t), y(t))$ is *periodic* if $(x(0), y(0)) = (x(T), y(T))$ for some $T > 0$. This solution defines then a closed curve in the phase plane, which is called a *limit cycle*.

A periodic solution is *stable* if the solutions that begin close to the limit cycle remain close to the limit cycle, while it is called *unstable* otherwise.

The pendulum. Consider the equation describing the movement of a pendulum

$$x'' = -\sin x.$$

This can be written as a system of autonomous differential equations of order 1 by:

$$\begin{aligned} x' &= v, \\ v' &= -\sin x. \end{aligned}$$

The equilibrium points are given by $v = 0$ and $\sin x = 0$. This means that $x = k\pi$, for every k. Let us study the stability.

The Jacobian of the system is

$$\begin{pmatrix} 0 & 1 \\ -\cos x & 0 \end{pmatrix}.$$

The eigenvalues of this matrix are $\lambda = \pm\sqrt{-\cos x}$. For the points of the form $x = k\pi$, we have that $\cos x = \pm 1$.

If $k = 0$, then the eigenvalues are $\lambda = \pm i$. The real part is zero, and hence this equilibrium point is critical. For $k = \pi$, the eigenvalues are $\lambda = \pm 1$, and hence the equilibrium point is unstable.

This is not very realistic. The intuition tells us that the point with $k = 0$, should be stable. The problem is that we are not taking into account the friction effects. For that, we should consider the differential equation to be

$$x'' = -\sin x - bx', \quad b > 0.$$

This can be written as a system of autonomous differential equations of order 1 by:

$$\begin{aligned} x' &= v, \\ v' &= -\sin x - bv. \end{aligned}$$

The equilibrium points are the same as before: $v = 0$ and $\sin x = 0$. The Jacobian of the system is now

$$\begin{pmatrix} 0 & 1 \\ -\cos x & -b \end{pmatrix}.$$

For $k = 0$, $\cos x = 1$ and the characteristic polynomial is $\lambda^2 + b\lambda + 1 = 0$. The solutions of this equation are

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4}}{2}.$$

If $b < 2$, $\sqrt{b^2 - 4}$ is imaginary. Hence, the real part of the eigenvalues is always negative. If $b = 2$, we get a negative real number. If $b > 2$, then $sqrtb^2 - 4$ is real, but for sure $b > \sqrt{b^2 - 4}$. Therefore, the eigenvalues are also negative.

Whatever the friction coefficient is, we see that the point $x = 0$ is a stable equilibrium point, as we could guess from experience.

Finally, let us see how are the trajectories in the non-friction case. We have to solve the implicit equation

$$\frac{dv}{dx} = \frac{-\sin x}{v},$$

which is

$$vdv = -\sin x dx.$$

Integrating at both sides from x_0 to x, we get:

$$v^2 = v_0^2 + 2(\cos x - \cos x_0).$$

Since the left hand side is positive, so must be the right hand side. This tells us that not all angles x are possible after an initial position x_0 is given.

Indeed, assume that the initial velocity is 0, that is, we start the pendulum from some angle x_0 and give no velocity. Then, the angles obtained as the time passes $x(t)$ satisfy that $\cos x - \cos x_0$ is positive, that is,

$$\cos x \geq \cos x_0.$$

This is satisfied when $x \in [x_0, -x_0]$, which is exactly what we know: the pendulum will not go up the angles $\pm x_0$.

8 Numerical analysis

8.1 Numerical methods for ordinary differential equations

Assume we are given an autonomous ODE of order 1:

$$y' = f(y),$$

with initial condition $y(0) = y_0$. Too often one is not able to find the solution function $y(t)$ satisfying the equation with the given initial condition.

In that case, one is left to use numerical methods that find an approximation of the solution. This approximation does not consist of an explicit function, but of a collection of values $y_0, y_1, y_2, \ldots$ approximating the function at times $t_0, t_1, t_2, \ldots$. That is, y_i is an approximation of the value $y(t_i)$.

The times $t_0, t_1, t_2, \ldots$ are called the *mesh points* and are usually taken to be of the type $t_1 = h + t_0, t_2 = 2h + t_0, \ldots$, with h called the *step size*.

These numerical methods are called *integration method of ODE's*. Interpolation methods allow one to obtain an approximate solution from the collection of approximate values $y_0, y_1, y_2, \ldots$.

In general, the smaller the time interval is, the better the approximation is, but the longer time is required to find the approximation.

We explain first of all the simplest method to approximate the solution of an ODE, the *Euler's method*. Then, we use one of the implemented methods in Octave and study the results.

Euler's method. Let $t_i = ih + t_0$ be the mesh points where we want to approximate the solution and h the step size.

Euler's method consists of approximating the solution function $y(t)$ by its first Taylor approximation at a value $t = a$:

$$y(t) = a + y'(a)(t - a) + \text{ error term.}$$

The differential equation $y' = f(y)$ tells us that $y'(a) = f(y(a))$, and hence

$$y(t) = a + f(a)(t - a) + \text{ error term.}$$

So, starting with an initial condition $y_0 = y(0)$, we construct the approximation values y_{i+1} as follows:

$$y_{i+1} = y_i + f(t_i)h.$$

That is, we assume that at a point y_i, the solution function is given by the first order Taylor polynomial at t_i:

$$P_{1,t_i}(t) = y_i + f(t_i)(t - t_i),$$

and consider the value of this polynomial at the next time $t_{i+1} = t_i + h$.

Lsode of Octave. In Octave there are a couple of methods for solving ODE's and systems of ODE's, namely, the "lsode" function and the "dassl" function.

We will use now the "lsode" function to understand the approximation of solutions.

We start with the ODE

$$x' = x.$$

We know that the general solution is given by $x(t) = Ce^t$.

To enter a differential equation in Octave, use the following:

```
> function xdot=f(x,t)
> xdot=x;
> endfunction
```

We need to fix now the initial condition that we will use, (in this example, $x(0) = 1$ so that $C = 1$):

```
> x0=1;
```

Finally, we store in a vector t the mesh points:

```
> t=linspace(0,6,1000);
```

This creates a vector of 1000 elements equispaced from 0 to 6.

To solve the ODE with initial condition $x(0) = 1$, with 1000 mesh points between 0 and 6, we do:

```
> z=lsode("f",x0,t);
```

The variable z has now the approximate values of $x(t)$ at the mesh points we chose.

To plot them, we do

```
> plot(t,z)
```

We can plot in the same graphic the solution that we know is the real one: $x(t) = e^t$. For that, type:

```
> plot(t,z,"@4",t,e.^t)
```

The "@4" is an argument to define the style of the dots and the color. You will find the options of the plot command by typing

```
> help plot
```

Exercise 1: Obtain the approximation of the same ODE with fewer mesh points:

(i) With 100 points between 0 and 6: store the points in the variable t2.

(ii) With 25 points between 0 and 6: store the points in the variable t3.

(iii) With 7 points between 0 and 6: store the points in the variable t4.

Store the approximated values of the solution at the variables $z2, z3, z4$ respectively.

Observe that $t(334) = t2(34) = t3(9) = t4(3) = 2$, that is, for those mesh points we have the approximate solution at 2. In the plot, it seems that the approximation is exact. But, is that true? Compute the differences between the real value $x(2) = e^2$ with the obtained approximations $z(334), z2(34), z3(9), z4(3)$.

Question 1: Which is the best approximation? Why is that? Do you think it was worth to consider 1000 mesh points?

Let us now find approximations of the solution of a system of ordinary differential equations. We consider the example seen in class, describing the dynamics of a pendulum:

$$\begin{aligned} x' &= v \\ v' &= -\sin(x). \end{aligned}$$

We first enter the system in Octave. The vector of differentials (x', y') is called "xdot". The vector of solutions (x, y) is called "x":

```
> function xdot=g(x,t)
> xdot=zeros(2,1);        (to initialize xdot as a vector of two coordinates)
> xdot(1)=x(2);
> xdot(2)=-sin(x(1));
> endfunction
```

We initialize the system at $(x, y) = (1, 0)$:

```
> x0=[1;0];
```

And define 100 mesh points from 0 to 10:

```
> t=linspace(0,10,100);
```

We solve the system:

```
> z=lsode("g",x0,t);
```

Now z is a matrix of dimensions 100×2. The columns are the values of the solutions x, y and each row corresponds to one mesh point. You can see it by typing

```
> z
```

in the command line.

If we plot the values of x against the values of y, we obtain the approximation of the trajectory that goes through the point $(1, 0)$:

```
> plot(z(:,1),z(:,2))
```

We see that we obtain a circle, centered at the origin. Observe that the range of x is $[-1, 1]$. Recall that we saw that the values of the angles, given an initial angle x_0 and velocity 0, where ranging between $-x_0$ and x_0.

Now we can also see how the solutions $x(t)$ and $y(t)$ look like. To plot them in the same graphic do:

```
> plot(t,z(:,1),"4",t,z(:,2))
```

Try to see what you get with other initial conditions.

Exercise 2: Consider the system describing the movement of a spring:

$$\begin{aligned} x' &= y, \\ y' &= -x. \end{aligned}$$

Find out with Octave an approximation of the trajectories at $(1, 0)$ and at $(3, 4)$, by considering 100 mesh points between 0 and 10. Plot also the solution functions x, y in the same graphic.

Question 2: What is the shape of the trajectories you found? (be careful: check the ranges of x and y). How are the solution functions? Could you guess, from the plot and from the system a particular solution of the system of differential equations?

Exercise (optional): Consider the system

$$\begin{aligned} x' &= y, \\ y' &= -2x. \end{aligned}$$

We saw that the trajectories of the system are ellipses centered in the origin. Check this fact computationally.

8.2 Optimization

One of the most common mathematical problems that one encounters is the one of finding the minimum or maximum of a function of one or more variables.

For example, the concentration of a drug after its administration can be modelled by the differential equation:

$$\frac{dc(t)}{dt} = -\frac{c(t)}{\tau},$$

where $c(t)$ denotes the concentration at time t and τ is a constant. The solution of this differential equation is given by

$$c(t) = c_0 e^{-t/\tau}.$$

Imagine we have data $c_0, c_1, \ldots, c_N$ of the concentrations of the drug at times $t_0, t_1, \ldots, t_N$. We would like to guess the value of τ from the experimental data obtained. In order to do that, a measure of how "close" the data from the theoretical values are. A function measuring this is called a *cost function*.

For instance, in our problem, we could consider as a cost function the euclidian distance between the observed values $(c_0, \ldots, c_N)$ and the theoretical values $(c_\tau(t_0), \ldots, c_\tau(t_N))$ given by each τ. Then, we have the cost function:

$$f(\tau) = \sum_{i=0}^{N} (c_\tau(t_i) - c_i)^2 = \sum_{i=0}^{N} (c_0 e^{-t_i/\tau} - c_i)^2.$$

The minimum of this function gives us an estimate of the parameter τ.

This example belongs to the family of optimization problems called parameter estimation.

Other optimization problems found in Bioinformatics are for example the determination of the structure of a protein, by finding the configuration that minimizes the free energy, or the docking of proteins.

Constrained and unconstrained optimization problems. The problem of finding the minimum or maximum of a function $f(x_1, \ldots, x_n)$ can be constrained or unconstrained.

The problem is *constrained* if there are restrictions on the region where we want to find the minimum or maximum. The constrains are given in the form

$$\begin{aligned} g_j(x_1, \ldots, x_n) &\geq 0, \qquad \text{for } j = 1, \ldots, m \\ h_k(x_1, \ldots, x_n) &= 0, \qquad \text{for } k = 1, \ldots, k. \end{aligned}$$

The first kind of constrains are called *inequality constrains*, while the second type are called *equality constrains*.

If there are no constrains, the optimization problem is said to be *unconstrained*.

Example 8.1. The optimization problem:

$$\text{minimize } f(x_1, \ldots, x_n)$$

is unconstrained while the optimization problem

$$\begin{gathered} \text{minimize } f(x_1, \ldots, x_n) \\ \text{subject to } x_i \geq 0 \text{ and } x_1 + \cdots + x_n = 0 \end{gathered}$$

is constrained, with n inequality constrains and one equality constrain.

Numerical optimization. Finding the minimum or maximum of a function by analytical methods is not always possible. In order to overcome this trouble, numerical methods are used. These methods are based on an initial guessing or random choice of where the minimum or maximum is, and then an iterative method is started that should approximate to the optimal value, as more iterations are performed.

If the method finds the optimal value, we say that the method *converged* to the optimal value. The number of iterations and the computations to be done at each process determine the speed of the process.

One should fix a stopping criterium, that determines when we are happy enough with the found value. Usually, the stopping criterium is based on the distance between two consecutive iteration points.

Local and global methods. There are several numerical methods to find an approximation of the minimum or maximum of a function. These methods can be divided in two groups:

(i) *Local methods*: that find a minimum or maximum which is closer to our starting point.

(ii) *Global methods*: that aim to find the global minimum or maximum of the function. They usually start with several random starting points.

Local methods are faster than the global ones, but we take the risk to find the non-wanted solution. Global methods are usually slow. Among others, the *genetic algorithms* are global methods.

We explain here the gradient descent method (or steepest descent) and the Newton's method. These are local method for unconstrained problems. When one knows a priori that the function has only one minimum, then it is suitable. If not, an inspection of where the minimum could be is necessary.

Gradient descent - A local unconstrained optimization method. Consider a differentiable function of one variable $f(x)$ and assume we want to find a minimum.

Consider a point x_0. Then, we make the following observation:

(i) If x_0 is at the left of the closest minimum, then the derivative of f at x_0 is negative and to approach the minimum we have to increase x_0.

(ii) If x_0 is at the right of the closest minimum, then the derivative of f at x_0 is positive and to approach the minimum we have to decrease x_0.

This is the basis of the gradient descent method: given a point x_0, to approach the closest minimum we should move towards the direction given by minus the differential of x_0.

So, we chose an initial point x_0. Then, we construct iteratively points $x_1, x_2, \dots$ by

$$x_{i+1} = x_i - \gamma f'(x_i),$$

where γ is a small number fixed by the user.

In this way we construct a sequence of points that approach the minimum. We can stop when the differences between two consecutive points is small enough.

Let us see it with an example. Consider the function $f(x) = x^2 + x$. Solving the problem analytically, we see that the minimum is located at $x = 0.5$. Let us see it computationally.

We make a while loop in octave by doing:

```
> a=0;
> b=1;
> gamma=0.01;
> while(abs(a-b)> 10^-4)
> b=a;
> a=a-gamma*(2*a+1);
> endwhile
```

For future modifications, you have the loop saved in a file "gradientdescent.m". To run it, just type

```
> gradientdescent
```

Just make sure that the file is in the working directory.

The value 10^{-4} gives the stopping criterium. With the tolerance to this value, we do not find the exact minimum. Try changing this value to smaller numbers and see if you get the exact minimum.

We can also play in changing the speed γ. See what happens when you do that.

Exercise 3: Find out with the gradient descent method the minimum of the function $g(x) = -e^{\sin(x)} + x^2$, in the interval $(-10, 10)$. It is not easy to find the minimum analytically.

Use an initial plot to figure out where the minimum is and start the gradient search from an approximation. You need to compute the derivative of g to perform the gradient descent.

Insert a counter in the loop (that is, a variable "i" that returns the number of iterations of the loop). Vary the speed γ and the stopping criterium and observe the change in the number of iterations.

Start also with different initial values and observe the convergence.

Question 3: Where is the minimum? Explain the different initial values, gamma and tolerance that you used. Which one gives the best option?

Newton's method. One of the problems with the gradient descent method is the choice of the value of γ. The Newton method changes the value of γ at each step, by the inverse of the second derivative of the function (if it exists).

So, consider a function $f(x)$ twice differentiable. Then, the iteration in the Newton's method is given by

$$x_{i+1} = x_i - \frac{f'(x_i)}{f''(x_i)}.$$

As an example, we compute the minimum of the function $f(x) = (x-4)^2 + x$. For that, use the file "newtonmethod.m".

Exercise 4: Modify the code in order to find the minimum of the function $g(x) = -e^{\sin(x)} + x^2$, in the interval $(-10, 10)$. Count also the iterations used.

Question 4: In this case, and for the same initial value, which method needs fewer iterations, the Newton's method or the gradient descent (you can try different values of γ)?

Exercise 5: Find the minimum of the function $f(x) = x^4 - x^3 - 2x^2$ using the Newton's method.

Question 5: Which is the approximate minimum of the function $f(x) = x^4 - x^3 - 2x^2$ that you found?

Gradient descent and Newton's method in more than one variables. The gradient descent method and the Newton's method also work for functions in more than one variable, by considering the partial derivatives and the second order partial derivatives.

For example, consider the function $f(x, y) = x^2 + y^2 - xy$ and the problem of finding the minimum.

We apply gradient descent to each variable. That is, given an initial point (x_0, y_0), suppose we have constructed iterative points up to (x_i, y_i). We construct the next point by

$$(x_{i+1}, y_{i+1}) = (x_i, y_i) - \gamma(\partial f/\partial x, \partial f/\partial y).$$

We have

$$\begin{aligned} \partial f/\partial x &= 2x - y, \\ \partial f/\partial y &= 2y - x. \end{aligned}$$

The loop is written in octave code in the file "gradientdescent2var.m".

For the Newton's method, one has to consider the so-called Hessian matrix of a function f:

$$H(x, y) = \begin{pmatrix} \frac{\partial^2 f}{\partial^2 x} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial^2 y} \end{pmatrix}.$$

We denote by $H(x_i, y_i)$ the evaluation of the Hessian at the point (x_i, y_i).
The iteration process is written as follows:

$$(x_{i+1}, y_{i+1}) = (x_i, y_i) - H(x_i, y_i)^{-1} \begin{pmatrix} \partial f/\partial x \\ \partial f/\partial y \end{pmatrix}.$$

Here $H(x_i, y_i)^{-1}$ is the inverse of the Hessian matrix. Observe that a necessary condition to apply the method is that the matrix is invertible.

Question 6: Write the octave code to find the minimum of the function $f(x, y) = x^2 + y^2 - xy$ by the Newton's method.

Octave - Successive quadratic programming optimization. In Octave there is an implemented optimization method called "successive quadratic programming". It is based in finding the minimum of successive second order approximations of our function.