Linear Algebra

Pascal Gourdel[1]

January 17, 2019

[1]Some material in this chapter (but not all of it) is from *Linear Algebra*, a very complete book by Jim Hefferon, from Saint Michael's College, Vermont, USA. His book is available on the web for free and is a fruitful reading for students.

Contents

Chapter 1

Linear computations in $\mathbb{R}^n$

1.1 Linear systems and Gauss' method

The Gauss' method to solve a linear system consists in several applications of Gaussian operations, until reaching an equivalent system in echelon form, that is very easy to solve. Let us define this.

A *linear equation* in variables $x_1, x_2, \ldots, x_n$ has the form

$$a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n = b \tag{1.1}$$

where the numbers $a_1, \ldots, a_n \in \mathbb{R}$ are the equation's *coefficients* and $b \in \mathbb{R}$ is the *constant*. An n-tuple $(s_1, s_2, \ldots, s_n) \in \mathbb{R}^n$ is a *solution* of, or *satisfies* that equation if substituting the numbers s_1, $\ldots$, s_n for the variables gives a true statement: $a_1s_1 + a_2s_2 + \ldots + a_ns_n = b$. Note that $0 = 1$ is linear equation, that has no solutions. Also, $0 = 0$ is a linear equation and every n-tuple $(s_1, s_2, \ldots, s_n) \in \mathbb{R}^n$ is a solution of the equation $0 = 0$.

A *system of linear equations* is a sequence of linear equations in the same set of variables:

$$\begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n = b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n = b_2 \\ \qquad\qquad\qquad\qquad\qquad\qquad \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n = b_m \end{cases} \tag{1.2}$$

It has the solution $(s_1, s_2, \ldots, s_n)$ if that n-tuple is a solution of all of the equations in the system.

Theorem 1 (Gauss operations) *If a linear system is changed to another by one of these operations*

1. *an equation is swapped with another;*
2. *an equation has both sides multiplied by a nonzero constant;*
3. *an equation is replaced by the sum of itself and a multiple of another*

then the two systems have the same set of solutions.

The three operations from Theorem 1 are the *elementary reduction operations*, or *row operations*, or *Gaussian operations*. They are *swapping*, *multiplying by a scalar* (also called *rescaling*) and *pivoting*. Before proving the theorem, the following lemma is useful:

Theorem 2 *The Gaussian operations are reversible. That is, if S is changed to S' by a Gaussian operation, then there exists a Gaussian operation that changes S' to S.*

Proof 1 : If S' is obtained from S by swapping two rows, then the same swapping applied to S' gives S again. If S' is obtained from S by multiplying a row by a scalar $\lambda \neq 0$, then multiplying the same row of S' by $1/\lambda$ gives S again. If S' is obtained from S by replacing the row i by itself plus the row j multiplied by a scalar λ, replacing the row i of S' by itself minus the row j of S' multiplied by λ gives S again.

Proof of Theorem 1: Let S' be a system on variables $(x_1, \ldots, x_n)$ obtained from S by applying a gaussian operation. We claim that if $(s_1, \ldots, s_n)$ is a solution of S, then it is a solution of S'. If S' is obtained from S by swapping two rows, this is clear. If S' is obtained from S by multiplying a row by a scalar, this is clear also. If S' is obtained by a pivoting, this is clear again. So our claim is clear. We still have to check the converse. But the converse can be proved by the same way since by Theorem 2, S is obtained from S' by a Gaussian operation. This proves Theorem 1.

Let us now describe formally the Gauss algorithm.

In each row of a linear system, the first variable with a nonzero coefficient is the row's *leading variable*. A system is in *echelon form* if each leading variable is to the right of the leading variable in the row above it (except for the leading variable in the first row). The non-leading variables in an echelon-form linear system are *free variables*.

Theorem 3 (Solving systems in echelon form)

- *A system in echelon form has at least a solution if and only if it contains no equation stating that 0 equals a non-zero number.*
- *Moreover, if a system in echelon form has at least one solution, then for every assignment of values to its free variables, it has exactly one solution.*
- *In particular, a system in echelon form has a unique solution if and only if it has no free variable and no equation stating that 0 equals a non-zero number.*

Proof 2 : The last claim follows trivially from the two first claims. Moreover, if the rows of the system S that we consider are all $0 = 0$, then the theorem is clear. Also, if the system has at least one row $0 = b$ where $b \neq 0$ then clearly S has no solution and the theorem holds. So we assume that S has at least one row that is not $0 = 0$, and has no row $0 = b$ with $b \neq 0$. Let us prove the two first claim of the theorem by induction on the number of rows of the system S.

Suppose first that S consists in a single row. Since this row is not $0 = b$, it is has a leading variable x_j, $1 \leq j \leq n$. So the first claim of the theorem holds. Because the system has at least one solution: give value 0 to every non-leading variable, and $b_1/a_{1,j}$ to x_j. For the second claim, let us assign values to every free variable. The resulting row may be written $a_{1,j}x_j = c$ where c is a constant. This has a unique solution: $c/a_{1,j}$.

Suppose that S has at least two rows. Since it has a row that is not $0 = b$, let us consider the last row i of S that has a leading variable x_j. Note that the variables $x_{j'}$, $j \leq j' \leq n$ are free variables of the system obtained from S by deleting Row i. Hence, assigning values to the free variable of S forces x_i to get a value by the paragraph above, and forces all the other leading variables by the induction hypothesis, providing a unique solution.

A row $a_{i,1}x_1 + \cdots + a_{i,n}x_n$ of a system is *leading* if either it has no leading variable and $a_{i',j'} = 0$ whenever $i' > i$ and $1 \leq j \leq n$, or its leading variable x_j is such that $a_{i',j'} = 0$ whenever $i' > i$ and $j' \leq j$. Note that a system is in echelon form if and only if every row is leading.

The *Gauss algorithm* uses the following strategy in order to turn a linear system into an equivalent system in echelon form: increase by one the number of consecutive leading rows at the beginning of a system, until reaching a system in echelon form. If not every row is leading, let i be the smallest index of a non-leading row. Let $i' \geq i$ be the indice of a row with a smallest leading variable indice. Such an indice i' exists because row i is not leading. Let us swap rows i and i'. Our new system has $i-1$ leading rows followed by one potentially non-leading row $a_{i,1}x_1 + \cdots + a_{i,n}x_n$ with a leading variable x_j. This i'th row is called the *pivot*. By the choice of i, j, $a_{i,j} \neq 0$, $a_{i,j'} = 0$ whenever $1 \leq j' < j$, and $a_{i',j'} = 0$ whenever $1 \leq j' < j$ and $i' > i$. For every $i < i' \leq m$ let us replace the row i' by the sum of itself and $-a_{i',j}/a_{i,j}$ multiplied by the pivot row. In this new system, the rows $1, \ldots, i$ are leading, so we increased by one the number of leading rows at the beginning of the system. Applying this fundamental step as long as there exist non-leading rows, we obtain:

Theorem 4 *For every linear system there is a finite sequence of gaussian operation that leads to an equivalent linear system in echelon form.*

1.2 How to solve and present the solution of a system

Here, we give examples that show how to deal concretely with various kind of linear systems.

Here is a system:

$$\begin{cases} 2x & +y & -z & +3t = 1 \\ 4x & +2y & -z & +4t = 5 \\ 2x & +y & & +t = 4 \end{cases}$$

We apply several step of Gaussian operations to obtain an equivalent system in echelon form.

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

The last row, that obtain by applying the Gaussian algorithm may be forgotten. This system has no "silly row" such as $0 = 1$, it has two leading

variables x and y, two free variables y and t, so it has infinitely many solutions according to Theorem 3. A good way to actually find the solution of the system is now to push the free variables on the right side of every equality in the system:

$$\begin{cases} 2x - z = 1 - y - 3t \\ \quad z \;\; = 3 \quad\quad + 2t \end{cases}$$

We substitute the value of z in the first row so that we get in every row a unique leading variable of the system and free variables:

$$\begin{cases} 2x - (3 + 2t) = 1 - y - 3t \\ \quad\quad z \quad\quad = 3 \quad\quad + 2t \end{cases}$$

Now it easy to turn the system into a system where every row is of the form $A = B$ where A is a leading variable and B is some expression where only free variables are involved.

$$\begin{cases} x = 2 - y/2 - t/2 \\ z = 3 \quad\quad\quad + 2t \end{cases}$$

The system above is a first way to present the solution. Indeed, every leading variable is a function of the free variables. This should be understood as: "any solution of the system can be made as follows: choose any values for the free variables, and then use the equations to compute the values of the leading variables". This way of presenting the system has a default: the lake of symmetry. In fact, by carrying the computations differently, it may happen that the set of free variables changes. This is why we sometimes prefer to present the solution as a *set of vectors*, but this will be done in the next section.

Chapter 2

Matrices and vectors of $\mathbb{R}^n$

The product of a $m \times n$ matrix A with a vector $x \in \mathbb{R}^n$ is a notion that allows to write linear systems in a compact form.

2.1 Definitions, rules

2.1.1 definition

An $m \times n$ *matrix* is a rectangular array of numbers with m *rows* and n *columns*. When $n = m$ the matrix is *square*. Each number in the matrix is an *entry*.

Usually, matrices will be denoted by upper-case letter, such as A, and their entries will be denoted by lower-case letters, with two subscripts: the first one corresponding to the row, the second one to the column. For instance, $a_{1,2}$ denotes the entry on the first line, second column, of A. We write $A = (a_{i,j})_{1 \leq i \leq m, 1 \leq j \leq n}$ or sometimes $A = (a_{i,j})$ if the number of rows and columns are clear from the context.

A *vector* (or *column vector*) is a matrix with a single column. A matrix with a single row is a *row vector*. The entries of a vector are its *components*.

Usually, vectors are denoted by lower-case letters. A vector whose components are all 0 is a *zero-vector* simply denoted by 0. By $\mathbb{R}^n$, we mean the set of every vectors with n real components. Thus, for us, an element of $\mathbb{R}^n$ is a column of n numbers. Note that in the chapter Logic and Sets, $\mathbb{R}^n$ denotes the sets of every n-tuples of reals, that are usually denoted by a row of n numbers. Distinguishing between rows and columns might seem spurious but in linear algebra, it may matter as we will see later.

For any matrix A, the *transpose* of A, written A^T, is the matrix whose columns are the rows of A. More precisely, if $A = (a_{i,j})_{1 \leq i \leq m, 1 \leq j \leq n}$ then $A = (a_{j,i})_{1 \leq j \leq n, 1 \leq i \leq m}$.

Note that the transpose of a vector is a row-vector, and conversely. Transpose is a fundamental tool in linear algebra but, by now, we use it only as a typographic trick: since a column is uncomfortable to typeset, we will sometimes write $v = (x_1, \ldots, x_n)^T$ instead of the equivalent statement $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

For any vectors $x, y \in \mathbb{R}^n$, the *sum* of x and y is the vector $z = x + y$ defined by $z_i = x_i + y_i$ for every $1 \leq i \leq n$. For any vector $x \in \mathbb{R}^n$ and any real number $t \in \mathbb{R}$, the *product* of x by t is the vector $z = tx$ defined by $z_i = tx_i$ for every $1 \leq i \leq n$.

We are now able to present the solution of the system of the section above as the set S of these vectors $(x, y, z, t)^T$ that satisfy the system.

$$\begin{aligned} S &= \{(2 - y/2 - t/2, y, 3 + 2t, t)^T \text{ such that } y, t \in \mathbb{R}\} \\ &= \{(2, 0, 3, 0)^T + y(-1/2, 1, 0, 0)^T + t(-1/2, 0, 2, 1)^T \text{ such that } y, t \in \mathbb{R}\} \end{aligned}$$

2.1.2 Product of a matrix by a vector

We will see later that matrices give a common way to write (surprisingly ?) many maps from $\mathbb{R}^n$ to $\mathbb{R}^n$ that have interesting combinatorial or geometric meanings.

For any $m \times n$ matrix A and any vector x of $\mathbb{R}^n$, the product of A by x is the vector of $\mathbb{R}^m$ $y = Ax$ defined by $y_i = a_{i,1}x_1 + \cdots + a_{i,n}x_n$ for every $1 \leq i \leq m$.

Theorem 5 *Let A be an $m \times n$ matrix, $x^1, \ldots, x^p \in \mathbb{R}^n$ be vectors and $\lambda_1, \ldots, \lambda_p$ be real numbers. We have $A(\lambda_1 x^1 + \cdots + \lambda_p x^p) = \lambda_1 Ax^1 + \cdots + \lambda_p Ax^p$.*

2.2 Applications

2.2.1 Gaussian operations in terms of matrices

Let $x_1, \ldots, x_k \in \mathbb{R}^n$ be vectors and $\lambda_1, \ldots, \lambda_k$ be real numbers. the vector $\lambda_1 x_1 + \cdots + \lambda_k x_k$ is the *linear combination* of $x_1, \ldots, x_k$ with coefficients $\lambda_1, \ldots, \lambda_k$.

Note that the columns of an $m \times n$ matrix A are n vectors of $\mathbb{R}^m$. If $x = (x_1, \ldots, x_n)^T$ then Ax is the linear combination of the columns of A with coefficients $x_1, \ldots, x_n$.

Gaussian row-operations applied to a matrix are: *swapping two rows*, *scaling a row* (that is multiplying it by a nonzero scalar) and *pivoting* (that is replacing a row by the sum of itself and a multiple of another). In each row of a matrix, the first nonzero entry is the row's *leading entry*. A matrix is in *echelon form* if each leading entry is to the right of the leading entry in the row above it (except for the leading variable in the first row). Applying Gaussian reduction to matrices instead of linear systems yields:

Theorem 6 *For every matrix, there is a sequence of Gaussian row-operations that leads to a matrix in echelon form.*

2.2.2 Linear systems revisited

Note that the linear system in Subsection 1.1 can now be written $Ax = b$ where $A = (a_{i,j})_{1 \le i \le m, 1 \le j \le n}$ and $b = (b_1, \ldots, b_m)^T$.

Theorem 7 *Let $S : Ax = b$ be a linear system where A is an $m \times n$ matrix and $b \in \mathbb{R}^m$. Then either:*

- *S has a unique solution;*
- *S has infinitely many solutions and there exist $k \ge 1$ and vectors $x_0, f_1, \ldots, f_k$ so that x is a solution if and only if there exist real numbers $\lambda_1, \ldots, \lambda_k$ satisfying $x = x_0 + \lambda_1 f_1 + \cdots + \lambda_k f_k$;*

  Moreover, k is number of free variables in the echelon system obtained from S after the Gaussian algorithm.
- *the system has no solution and there is a linear combination of its equations that yields $0 = 1$.*

Moreover, only one of these statements holds.

Proof 3 : It is clear that at most one of the statement holds since no solution can satisfies $0 = 1$. Let us prove that at least one of the statements holds. By Theorem 4 let us transform our system by a sequence of operations into a system S in echelon form. If S has a row stating $0 = b$ where $b \neq 0$ then by multiplying it by $1/b$, we obtain the desired row $0 = 1$. Otherwise, there is no row stating that $0 = b$ where $b \neq 0$. Then S as at least a solution by Theorem 3. If it has no free variable, this solution is unique again by Theorem 3, and if it has free variables, Theorem 3 says that there are infinitely many solutions since every choice of value brings a solution.

Theorem 8 *Let A be a $m \times n$ matrix and $b \in \mathbb{R}^m$ be a vector. If there is a vector $x_0 \in \mathbb{R}^n$ such that $Ax_0 = b$ then for any y*

$Ay = b$ if and only if $y = x_0 + x$ for some x such that $Ax = 0$.

$Ay = b$ if and only if $y - x_0$ is a solution of the system $Ax = 0$.

The theorem above says that to solve $Ax = b$, it suffices to have one solution x_0 of $Ax = b$. The other solutions are obtained by adding x_0 with the solutions of $Ax = 0$. A system $Ax = b$ is *homogeneous* if $b = 0$. The system $Ax = 0$ is the *homogeneous system associated to* $Ax = b$.

2.2.3 Map associated to a matrix

When A is an $m \times n$ matrix, we denote by f_A the map from $\mathbb{R}^n$ to $\mathbb{R}^n$ defined by $f_A(x) = Ax$. So, for any $x_1, \ldots, x_p \in \mathbb{R}^n$ vectors and $\lambda_1, \ldots, \lambda_p \in \mathbb{R}^n$ be vectors $f_A(\lambda_1 x_1 + \cdots + \lambda_p x_p) = \lambda_1 f_A(x_1) + \cdots + \lambda_p f_A(x_p)$ by Theorem 5.

Theorem 9 *Let A, B be two matrices. Then $f_A = f_B$ if and only if $A = B$.*

Proof 4 : If $A = B$ then clearly $f_A = f_B$. Conversely, if $f_A = f_B$ then A and B must have same dimension, say $m \times n$. For every $1 \leq i \leq n$, we denote by e_i the vector with only 0 components, except the i'th one that is 1. Then, $f_A(e_i)$ is easily seen to be the i'th column of A and $f_B(e_i)$ the i'th column of B. So, A and B must have the same columns, they must be equal.

The *identity matrix* I_n is the $n \times n$ matrix defined as follows: the (i, j) entry is 1 if $i = j$ and 0 otherwise. If x is a vector of $\mathbb{R}^n$, then $I_n x = x$. Hence $f_{I_n} = \mathrm{Id}$ where Id is the function identity. A square matrix A is *upper triangular* (resp. *lower*) if for every $i > j$ (resp. $i < j$), $a_{i,j} = 0$. A matrix that is either upper or lower triangular is simply *triangular*.

Theorem 10 *Let A be a $n \times n$ square matrix. The following properties are equivalent.*

1. *f_A is a bijection*
2. *for every $b \in \mathbb{R}^n$, $Ax = b$ has a unique solution*
3. *there exists a vector $b \in \mathbb{R}^n$ such that $Ax = b$ has a unique solution*
4. *$Ax = 0$ has a unique solution*
5. *There exists a sequence of row operations that applied to A leads to a triangular matrix with non-zero entries on its diagonal.*
6. *There exists a sequence of row operations that applied to A leads to the identity matrix.*

Proof 5 : (1) $\leftrightarrow$ (2) : since f_A is a bijection, for every $b \in \mathbb{R}^n$ there is a unique x so that $f_A(x) = b$. Since $f_A(x) = Ax$, this means exactly that $Ax = b$ has a unique solution.

(2) $\rightarrow$ (3) : clear.

(3) $\rightarrow$ (4) : clear by Theorem 8.

(4) $\rightarrow$ (5) : let us transform $Ax = 0$ into a system in echelon form by Gaussian operations. If every row has a leading variable, then since A is square, there is no free variable and for every i, x_i must be the leading i'th row of the system. So the same sequence of Gaussian operations applied to A lead to a triangular matrix with non-zero entries on its diagonal as claimed. Else, there is a row with no leading variable, so some there must exists at least one free variable. Hence, either the system has no solution or it has infinitely many solution, in both cases contradictory to our hypothesis.

(5) $\rightarrow$ (6) : example ...

(6) $\rightarrow$ (2) : Apply the sequence of Gaussian operations to $Ax = b$ leads to a system of the form $I_n x = c$, which has a unique solution c.

A matrix satisfying the properties of the theorem above is *non-singular*, and *singular* otherwise.

2.2.4 Operations on matrices

The *sum* of two same-sized matrices is their entry-by-entry sum. The *scalar multiple* of a matrix is the result of entry-by-entry scalar multiplication.

Since matrices represent functions, it is natural to ask which matrix corresponds to the composition of two functions. This leads us to the notion of product of two matrices.

The *matrix-multiplicative product* of the $m \times r$ matrix A and the $r \times n$ matrix B is the $m \times n$ matrix C, where

$$c_{i,j} = a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \ldots + a_{i,r}b_{r,j} \tag{2.1}$$

$$AB = \begin{pmatrix} & \vdots & \\ a_{i,1} & a_{i,2} & \ldots & a_{i,r} \\ & \vdots & \end{pmatrix} \begin{pmatrix} & b_{1,j} & \\ \ldots & b_{2,j} & \ldots \\ & \vdots & \\ & b_{r,j} & \end{pmatrix} = \begin{pmatrix} & \vdots & \\ \ldots & c_{i,j} & \ldots \\ & \vdots & \end{pmatrix} \tag{2.2}$$

Note that the row i of AB is the linear combination of the rows of B whose coefficients are the numbers in the row i of A. The column j of AB is the linear combination of the columns of A, whose coefficients are the numbers in the column j of B.

2.2.5 Trace and transpose

Theorem 11 *Let A be an $m \times l$ matrix and B be an $l \times n$ matrix. Then $(\lambda A + \mu B)^T = \lambda A^T + \mu B^T$ and $(AB)^T = B^T A^T$.*

The *trace* of a square matrix M is the sum denoted by $\mathrm{Tr}(M)$ of the entries on its diagonal.

Theorem 12 *Let A, B be $n \times n$ matrices. Then $\mathrm{Tr}\,(AB) = \mathrm{Tr}(BA)$.*

2.2.6 Linear geometry of $\mathbb{R}^n$

No definition or theorem of geometry is requested for this course. However, geometry is a natural way of "seeing" what is going on with linear algebra, and helps a lot to understand and remember the definitions and theorems.

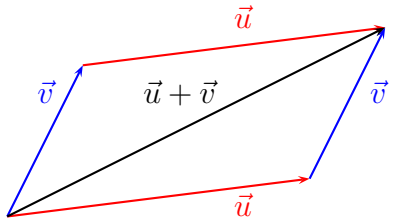

Figure 2.1: sum of two vectors

We assume that the following notions are known by the reader: coordinates of a point in 2 and 3 dimensional space, distance, angles, rotation, function sinus and cosinus. The objects and statements given in this subsection will not be defined rigorously nor proved.

Geometrically, what we call a vector is defined by a length and a direction. They may be represented by arrows. Note that two arrows of same length and same direction represent the same vector, even if they are drawn at different places. To obtain the sum of two vectors geometrically, draw the parallelogram formed by them. Their sum extends along the diagonal to the far corner. To obtain the coordinates of a vector geometrically, draw that vector starting at the origin. The coordinates of the vector are the coordinates of the point where the vector gets to.

The following result has to be known: the set of point of coordinate (x, y) in the plane $\mathbb{R}^2$ such that $ax + by = 0$ is a line and the vector $(a, b)^T$ is perpendicular to that line. Similarly, the set of points of coordinate (x, y, z) in the space $\mathbb{R}^3$ such that $ax + by + cz = 0$ is a plane and the vector $(a, b, c)^T$ is perpendicular to that plane.

2.2.7 Some matrices and vectors of interest

If $\lambda_1, \ldots, \lambda_n$ are real numbers, then the linear combination of the real numbers $x_1, \ldots, x_n$ with coefficients $\lambda_1, \ldots, \lambda_n$ equals $(\lambda_1, \ldots, \lambda_n)(x_1, \ldots, x_n)^T$. Hence the matrix $(\lambda_1, \ldots, \lambda_n)$ is the matrix associated to the linear combination with coefficients $\lambda_1, \ldots, \lambda_n$. The result is sometimes called the *dot product* of $(\lambda_1, \ldots, \lambda_n)$ and $(x_1, \ldots, x_n)^T$. Note that when A, B are matrices, then the i, j-th entry of the product AB is the $d$ot product of the i-th row and the j-th column.

A *permutation* of $\{1, \ldots n\}$ is a bijection from $\{1, \ldots, n\}$ to itself.

Let r be the rotation of angle θ centered at the origin in the plane. If the vector x is sent to the vector x by r, we put $y = r(x)$. We have:

$$y = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} x$$

This is why the matrix in the equation above is *the matrix associated to the rotation of angle* θ. To see this, we assume that we know by geometry that $r(\lambda x + \mu y) = \lambda r(x) + \mu r(y)$. Let x be the vector of coordinates $(x_1, x_2)^T$ and $e_1 = (1, 0)^T$, $e_2 = (0, 1)^T$. By geometry, we know that $r(e_1) = (\cos(\theta), \sin(\theta))^T$ and that $r(e_2) = (-\sin(\theta), \cos(\theta))^T$. Hence, $y = r(x) = r(x_1e_1+x_2e_2) = x_1r(e_1)+x_2r(e_2) = x_1(\cos(\theta), \sin(\theta))^T+x_2(-\sin(\theta), \cos(\theta))^T = (\cos(\theta)x_1 - \sin(\theta)x_2, \sin(\theta)x_1 + \cos(\theta)x_2)^T$.

There are matrices associated to Gaussian operations. If n is an integer, if $1 \leq i, j \leq n$ are integers and λ is a real number, then we put:

- if $i \neq j$, $P_{n,i,j,\lambda}$ is the $n \times n$ matrix whose entries are all 0, except for the diagonal whose entries are 1, and the entry (i, j) that is λ;
- if $\lambda \neq 0$, $P_{n,i,i,\lambda}$ is the $n \times n$ matrix whose entries are all 0, except for the diagonal whose entries are 1, except for the entry (i, i) that is λ;
- $P_{n,i,j}$ is the $n \times n$ matrix obtained from I_n by swapping rows i and j

Note that all the matrices defined above are non-singular.

Theorem 13 *Let A be an $m \times n$ matrix. Then:*

- *the matrix obtained from A be swapping rows i, j is $P_{m,i,j}A$;*
- *the matrix obtained from A be multiplying row i by a scalar λ is $P_{m,i,i,\lambda}A$;*
- *the matrix obtained from A be replacing row i by itself plus λ times row j is $P_{m,i,j,\lambda}A$;*
- *the matrix obtained from A be swapping columns i, j is $AP_{n,i,j}$;*
- *the matrix obtained from A be multiplying column j by a scalar λ is $AP_{n,j,j,\lambda}$;*
- *the matrix obtained from A be replacing column j by itself plus λ times column i is $AP_{n,i,j,\lambda}$.*

2.2.8 Gaussian operations revisited

If M' is obtained from M by a sequence of Gaussian row-operations, then every row of M' is a linear combination of the rows of M. This can be seen by back-tracking the row-operations.

Theorem 14 *For every $(n+1) \times n$ matrix there is a linear combination (with at least one nonzero coefficient) of the rows that yields 0.*

Proof 6 : Apply Theorem 6 and note that in an $m \times n$ matrix in echelon form, at most n rows are nonzero.

Chapter 3

Vector space

3.1 Definition of a vector space

Vectors spaces may be seen as a generalization of $\mathbb{R}^n$ that preserves its properties with respect to linear combinations. Such a generalization is relevant since a lot of sets of interest (set of the solutions to some differential equations, sets of polynomial of degree at most n, ...) are vector spaces.

Classical notations of a vector use arrow ($\vec{u}$,$\overrightarrow{AB}$,...). It is not an obligation, just an useful convention in order to distinguish scalars (here real numbers) from vectors. When people are familiar with linear algebra, such a notation is unnecessary.

Definition 1 *A* vector space *(over $\mathbb{R}$) consists of a set V along with two operations '+' and '$\cdot$' subject to Conditions (1) to (10) below.*

When $\vec{x}, \vec{y} \in V$,

(1) their vector sum *$\vec{x} + \vec{y}$ is an element of V.*

(2) $\vec{x} + \vec{y} = \vec{y} + \vec{x}$ and

If $\vec{x}, \vec{y}, \vec{z} \in V$ then

(3) $(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$.

(4) There is a zero vector *$\vec{0} \in V$ such that $\vec{x} + \vec{0} = \vec{x}$ for all $\vec{x} \in V$.*

(5) Each $\vec{x} \in V$ has an additive inverse *$\vec{y} \in V$ such that $\vec{x} + \vec{y} = \vec{0}$.*

If r is a scalar *(element of $\mathbb{R}$), and $\vec{x} \in V$ then*

(6) each scalar multiple $r \cdot \vec{x}$ *is in V.*

If $r, s \in \mathbb{R}$ and $\vec{x}, \vec{y} \in V$ then

(7) $(r+s) \cdot \vec{x} = r \cdot \vec{x} + s \cdot \vec{x}$, *and*

(8) $r \cdot (\vec{x} + \vec{y}) = r \cdot \vec{x} + r \cdot \vec{y}$, *and*

(9) $(rs) \cdot \vec{x} = r \cdot (s \cdot \vec{x})$, *and*

(10) $1 \cdot \vec{x} = \vec{x}$.

Some classical rules can be deduced $0 \cdot \vec{x} = \vec{0}$, $r \cdot \vec{0} = \vec{0}$. Most of the times, the notation for the external law will be omitted : $\vec{u} + \vec{u} = 2 \cdot \vec{u} = 2\vec{u}$.

examples $\mathbb{R}^n$... , $\mathcal{F}(A, E)$ if E is a vector space.

3.2 Vector subspace

For any vector space, a *subspace* is a subset that is itself a vector space, under the inherited operations. Note that every vector space must contain $\vec{0}$. The vector space containing only $\vec{0}$ is the *trivial subspace*.

Theorem 15 *The intersection of subspaces of a vector space V is a subspace of V.*

Note that there is no similar theorem with the union: the union of two subspaces may fail to be a subspace.

Theorem 16 *For a nonempty subset S of a vector space V, under the inherited operations, the following are equivalent statements.*

- *S is a subspace of that vector space*
- *S is closed under linear combinations of pairs of vectors: for any vectors $\vec{s}_1, \vec{s}_2 \in S$ and scalars r_1, r_2 the vector $r_1\vec{s}_1 + r_2\vec{s}_2$ is in S*
- *S is closed under linear combinations of any number of vectors: for any vectors $\vec{s}_1, \ldots, \vec{s}_n \in S$ and scalars $r_1, \ldots, r_n$ the vector $r_1\vec{s}_1 + \cdots + r_n\vec{s}_n$ is in S.*

The *span* (or *linear closure*) of a nonempty subset K of a vector space V is the set of all linear combinations of vectors from K. No notation for the span is completely standard. It will be denoted here either by $\text{vect}(K)$ or $\text{span}(K)$. It correspond geometrically to the set of points, that can be reached if you are only allowed to use directions given by elements of K.

The span of the empty subset of a vector space is the trivial subspace while the following formula explicits the value of $\text{vect}(K)$ if the set K is nonempty.

$$\text{vect}(K) = \left\{ \vec{z} \in V \;\middle|\; \begin{array}{l} \text{there exists } c_1, \ldots, c_n \in \mathbb{R}, \\ \text{there exists } \vec{s}_1, \ldots, \vec{s}_n \in K, \\ \text{satisfying } \vec{z} = c_1\vec{s}_1 + \cdots + c_n\vec{s}_n \end{array} \right\} \tag{3.1}$$

An alternative definition is given by the fact that the span of K is the inclusion-wise minimal vector subspace of V that contains K: the important rules are

$$\text{vect}(K) = K \Leftrightarrow K \text{ is a vector space}$$

$$K \subset G \text{ where } G \text{ is a vector space} \Rightarrow \text{span}(k) \subset G$$

Proposition 1 *The span of a subset K of vector space V is the intersection of all the subspaces of V that contain K.*

A subset of a vector space E is a *spanning set* if $\text{span}(K) = E$.

3.2.1 Linear independence

Linear independence is a notion that formalizes the fact, already met when solving systems, that some vectors in a list are in some respect "unnecessary" to describe the set.

A subset of a vector space is *linearly independent* if none of its elements is a linear combination of the others. Otherwise it is *linearly dependent*.

Proposition 2 *A subset S of a vector space is linearly independent if and only if for any distinct $\vec{x}_1, \ldots, \vec{x}_n \in S$ the only decomposition of the null vector is the trivial one*

$$\lambda_1\vec{x}_1 + \ldots + \lambda_n\vec{x}_n = \vec{0}, \text{ when } \lambda_1, \ldots, \lambda_n \in \mathbb{R} \Rightarrow \lambda_1 = 0, \ldots, \lambda_n = 0$$

Proposition 3 *If $\vec{x}_1, \ldots, \vec{x}_n \in V$ are independent vectors, and if $\vec{y} \in V$ is a vector such that $\vec{x}_1, \ldots, \vec{x}_n, \vec{y}$ are not independent, then $\vec{y}$ is a linear combination of $\vec{x}_1, \ldots, \vec{x}_n$.*

The last proposition can be translated in terms of equations. Let us consider two linear systems:

$$(\mathcal{S}) \quad \begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n = b_1 \\ \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n = b_m \end{cases}$$

and

$$(\mathcal{S}_2) \quad \begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n = b_1 \\ \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n = b_m \\ c_1x_1 + c_2x_2 + \cdots + c_nx_n = d \end{cases}$$

If $(\mathcal{S})$ does not contain any redundant equation while the "extended" linear system $(\mathcal{S})_2$ contains a redundant equation then the "additional" equation can be decomposed as a linear combination of the m first equations.

3.2.2 Basis and dimension

It is easy to see that any subset of an independent set is also independent, and that any set containing a spanning set is also spanning. Hence, maximal independent sets and minimal spanning sets are worth investigating.

The following is called *Steinitz' exchange property*. It is an important step in order to define the concept of dimension

Lemma 1 *Let $n \geq 1$ be an integer. In a vector space, if $\vec{x}_1, \ldots, \vec{x}_n$ are independent vectors and if $\vec{y}_1, \ldots, \vec{y}_{n+1}$ are independent vectors, then there exits $1 \leq i \leq n+1$ such that $\vec{x}_1, \ldots, x_n, \vec{y}_i$ are independent vectors.*

Proof 7 *Suppose not. By Proposition 3, for every $1 \leq i \leq n+1$, $\vec{y}_i = a_{i,1}\vec{x}_i \cdots + a_{i,n}\vec{x}_n$. By Theorem 14, there is a linear combination (with at least one non-zero coefficient) of the rows of $(a_{i,j})$ that equals 0. Hence, the same linear combination applied to $\vec{y}_1, \ldots, \vec{y}_{n+1}$ equals $\vec{0}$, contradictory to the independence of $\vec{y}_1, \ldots, \vec{y}_{n+1}$.*

Proposition 4 *Let $n \geq 1$ be an integer. In a vector space, if $\vec{x}_1, \ldots, \vec{x}_n$ are independent vectors and if $\vec{x}_{n+1} \notin \text{vect}(\vec{x}_1, \ldots \vec{x}_n)$ then $\vec{x}_1, \ldots, \vec{x}_n, \vec{x}_{n+1}$ are independent vectors.*

Theorem 17 (Weak duality) *In a vector space, if $\vec{x}_1, \ldots, \vec{x}_n$ are independent vectors and if $\vec{y}_1, \ldots, \vec{y}_m$ are vectors that span that space, then $m \geq n$.*

Proof 8 *Let us choose $\{\vec{y}_{i_1}, \ldots, y_{\vec{i}_k}\}$ an independent set of maximum cardinality among the $\vec{y}_i$'s. By the maximality and by Proposition 3, every vector in $\{\vec{y}_1, \ldots, \vec{y}_m\}$ is a linear combination of $\vec{y}_{i_1}, \ldots, y_{\vec{i}_k}$, and since $\{\vec{y}_1, \ldots, \vec{y}_m\}$ spans the space, it follows that $\{\vec{y}_{i_1}, \ldots, \vec{y}_{i_k}\}$ spans the space. If $n < m$, then by Lemma 1, there exists an $\vec{x}_i$ such that $\vec{y}_{i_1}, \ldots, \vec{y}_{i_k}, \vec{x}_i$ are independent. So, $\vec{x}_i$ is not a linear combination of $\vec{y}_{i_1}, \ldots, \vec{y}_{i_k}$, a contradiction since $\vec{y}_{i_1}, \ldots, \vec{y}_{i_k}$ span the space.*

An independent family of vectors that span the space is a *basis*. For example, if $n \geq 1$ is an integer, let us define for every $1 \leq i \leq n$ the vector $\vec{e}_i$ of $\mathbb{R}^n$, whose components are all 0, except the i'th component that equals 1. Then $(\vec{e}_1, \ldots, \vec{e}_n)$ is a basis of $\mathbb{R}^n$, called the *canonical basis*. The vector set of every polynomial function of degree n has dimension $n+1$ and has a basis $1, x, x^2, \ldots$. Note that there is no canonical basis in a general vector space.

Corollary 1 *If $\mathcal{B} = (\vec{x}_1, \ldots, \vec{x}_n)$ and $\mathcal{C} = (\vec{y}_1, \ldots, \vec{y}_m)$ are basis then $n = m$.*

Proof 9 *Since $\vec{x}_1, \ldots, \vec{x}_n$ are independent vectors and since $\vec{y}_1, \ldots, \vec{y}_m$ are vectors that span that space, we can apply Theorem 17 in order to get $m \geq n$. But by a symmetric argument, we can remark that $\vec{y}_1, \ldots, \vec{y}_m$ are independent vectors and $\vec{x}_1, \ldots, \vec{x}_n$ are vectors that span that space, we can apply Theorem 17 in order to get the converse inequality.*

A vector space of *finite dimension* is a vector space that contains a finite spanning family.

The *dimension* of the space is the cardinal of a smallest spanning family.

Theorem 18 *Let $\vec{x}_1, \ldots, \vec{x}_n$ be vectors in a vector space V of finite dimension. The following properties are equivalent :*

- *$(\vec{x}_1, \ldots, \vec{x}_n)$ is a basis;*

- *V has dimension n and $\vec{x}_1, \ldots, \vec{x}_n$ are independent;*
- *V has dimension n and $\vec{x}_1, \ldots, \vec{x}_n$ span V;*
- *$\{\vec{x}_1, \ldots, \vec{x}_n\}$ is an inclusion-wise maximal independent set;*
- *$\{\vec{x}_1, \ldots, \vec{x}_n\}$ is an inclusion-wise minimal set that spans V;*
- *For every vector $\vec{y} \in V$ there is a unique linear combination of the $\vec{x}_i$'s that equals y.*

It follows that every vector space of finite dimension has at least one basis and that every basis has n vectors where n is the dimension of the space. It allows to consider the coordinates of any $\vec{y}$ since the decomposition is unique.

Theorem 19 *Every independent set of a vector space is included in some basis, and every spanning set contains a basis.*

3.2.3 Computing spans and dimensions

The two following theorems give a way to compute a basis of the span of a finite set of vectors of $\mathbb{R}^n$: write them as row vectors of matrix, use Gaussian operations on rows till reaching of matrix in echelon form. The non-zero rows of the matrix form a basis of the span, and their number is the dimension.

Theorem 20 *Let M be an $m \times n$ matrix.*

- *If M' is obtains from M by Gaussian operations on columns, then the subspace of $\mathbb{R}^m$ spanned by the columns of M is the same as the subspace of $\mathbb{R}^m$ spanned by the columns of M'.*
- *If M' is obtains from M by Gaussian operations on rows, then the subspace of $\mathbb{R}^n$ spanned by the rows of M is the same as the subspace of $\mathbb{R}^n$ spanned by the rows of M'.*

Theorem 21 *Let M be an $m \times n$ matrix in echelon form. Then:*

- *the columns of M that contain a leading entry form a basis of the subspace of $\mathbb{R}^m$ spanned by the columns of M;*

- *the rows of M that contain a leading entry (that are exactly the non-zero rows of M) form a basis of the subspace of $\mathbb{R}^n$ spanned by the row of M;*
- *the two spaces mentioned above have same dimension, that is the number non-zero rows of M.*

3.2.4 Sum, direct sum and complement

Definition 2 *Let E be a vector space, and V, W be vector subspaces of E. Let us consider the set*

$$V + W = \{\vec{z} \mid \text{there exists } \vec{x} \in V,\ \vec{y} \in W \text{ such that } \vec{z} = \vec{x} + \vec{y}\}.$$

This set is the sum *of W and W.*

Remark 1 *It is easy to check that $V + W$ is a vector space, that it contains both V and W since for example any $\vec{x} \in V$ can be decomposed $\vec{x} = \vec{x} + \vec{0}$ where $\vec{x} \in V$ and $\vec{0} \in W$. More precisely,*

$$V + W = span(V \cup W).$$

We can generalize

Definition 3 *Let E be a vector space, and $V_1, \ldots V_k$ be vector subspaces of E. Let us consider the set*

$$V_1 + \ldots + V_k = \{\vec{z} \mid \text{there exists } \vec{x}_k \in V_k, \text{ such that } \vec{z} = \vec{x}_1 + \ldots + \vec{x}_k\}.$$

Remark 2 *It is easy to check that $V_1 + \ldots + V_k$ is a vector space, that it contains all the V_i. More precisely,*

$$V_1 + \ldots + V_k = span(V_1 \cup \ldots \cup V_k).$$

For any $\vec{x} \in V_1 + \cdots + V_k$, there exists at least one decomposition $\vec{x} = \vec{x}_1 + \ldots + \vec{x}_k$ but in general the decomposition is not unique. This is why we will introduce the concept of direct sum.

Definition 4 *Let $V_1, \ldots, V_k$ be subspaces of a vector space E, we will say that they are in* direct sum *if for any $\vec{x} \in V_1 + \cdots + V_k$, the above decomposition is unique.*

We will denote by $W_1 \oplus \ldots \oplus W_k$ the set $W_1 + \ldots + W_k$ in order to keep in mind the information that the sum is direct. There exists an easy characterization (of particular interest when the length is equal to two).

Proposition 5 *Let $V_1, \ldots, V_k$ be subspaces of a vector space E. The sum $V_1 + \ldots + V_k$ is direct if and only if the null vector has a unique decomposition.*

Remark 3 *Let $V_1, \ldots, V_2$ be subspaces of a vector space E. The sum $V_1 + \ldots V_k$ is direct if and only if $V_1 \cap V_2 = \{\vec{0}\}$. Indeed, if $\vec{x} \in V_1 \cap V_2$, and $\vec{x} \neq \vec{0}$, then we can write several different decompositions (and many other)*

$$\vec{x} = \vec{0} + \vec{x} = \vec{x} + \vec{0} = (1/2)\vec{x} + (1/2)\vec{x}$$

An easy way to build a basis of $V_1 \oplus V_2$ is to the concatenate a basis of V_1 with a basis of V_2.

The case of two subspaces of E whose direct sum is E is so important that it is worth to introduce a particular terminology.

Definition 5 *When $E = V_1 \oplus V_2$, V_1 and V_2 are* complement *with respect to E.*

Theorem 22 *Every subspace has a complement. More precisely, if V is a vector subspace of E then there exits a subspace W such that $E = V \oplus W$.*

These results can be extended in a general theorem.

Theorem 23 *Let $V_1, \ldots, V_k$ be non-zero subspaces of a vector space E of finite dimension. The following conditions are equivalent:*

- $E = V_1 \oplus \cdots \oplus V_k$.
- $V_1 + \cdots + V_k = E$ *and* $\dim(V_1) + \cdots + \dim(V_k) = \dim(E)$;
- *for every $(\mathcal{B}_1, \ldots, \mathcal{B}_k)$ such that $\mathcal{B}_i$ is a basis of W_i ($1 \leq i \leq k$), the concatenation of the $\mathcal{B}_i$'s, $1 \leq i \leq k$, is a basis of E.*
- *for some $(\mathcal{B}_1, \ldots, \mathcal{B}_k)$ such that B_i is a basis of W_i ($1 \leq i \leq k$), the concatenation of the $\mathcal{B}_i$'s, $1 \leq i \leq k$, is a basis of E.*

3.3 Maps between spaces

3.3.1 Linear maps

A function f from a vector spaces V to a vector space W that preserves the operations of addition:

$$\text{if } \vec{x}_1,\ \vec{x}_2 \in V \text{ then } h(\vec{x}_1 + \vec{x}_2) = h(\vec{x}_1) + h(\vec{x}_2)$$

and scalar multiplication:

$$\text{if } \vec{x} \in V \text{ and } t \in \mathbb{R} \text{ then } h(t \cdot \vec{x}) = t \cdot h(\vec{x})$$

is a *linear map*. When $V = W$, that is when h is from V to itself, h is an *endomorphism*.

Note that a map f is linear if and only if for every $\vec{x}, \vec{y} \in V$ and $\lambda, \mu \in \mathbb{R}$ we have $f(\lambda\vec{x} + \mu\vec{y}) = \lambda f(\vec{x}) + \mu f(\vec{y})$.

Examples: $f(X) = AX$, $\phi(f) = df/dx$, $A \to ({}^T A)$.

Counter examples: trigo, x^2, $ax + 1$, e^x ...

Note that for every linear map $f \in \mathcal{L}(E, F)$, $f(\vec{0}_E) = \vec{0}_F$. In particular, a constant map is linear if and only if it is equal to $\vec{0}$.

Theorem 24 *A linear map is determined by its action on a basis. That is, if $(\vec{b}_1, \ldots, \vec{b}_n)$ is a basis of a vector space V and $\vec{c}_1, \ldots, \vec{c}_n$ are (perhaps not distinct) elements of a vector space W then there exists a linear map from V to W sending $\vec{b}_1$ to $\vec{c}_1$, ..., $\vec{b}_n$ to $\vec{c}_n$, and that linear map is unique.*

Proof 10 *Let us prove the existence. If $\vec{x} \in V$, then $\vec{x} = \lambda_1 b_1 + \cdots + \lambda_n \vec{b}_n$ since $(\vec{b}_1, \ldots, \vec{b}_n)$ is a basis. Since this decomposition is unique, it is well defined to consider a map defined f by: $f(\vec{x}) = \lambda_1\vec{c}_1 + \cdots + \lambda_n\vec{c}_n$. Checking that $f(\lambda\vec{x} + \mu\vec{y}) = \lambda f(\vec{x}) + \mu f(\vec{y})$ for every $x, y \in V$ and every λ, $\mu \in \mathbb{R}$ is a routine matter.*

The uniqueness is clear: since $\vec{c}_1 = f(\vec{b}_1), \ldots, \vec{c}_n = f(\vec{b}_n)$ and since we want f to be linear, we must have $f(\vec{x}) = \lambda_1 c_1 + \cdots + \lambda_n\vec{c}_n$.

3.3.2 Image and kernel

The *rangespace*, or *image* of a linear map $h\colon V \to W$ is

$$\text{Im}(h) = \{h(\vec{x}) \text{ such that } \vec{x} \in V\},$$

sometimes denoted $h(V)$. The dimension of the image is the map's *rank*. The *nullspace* or *kernel* of a linear map $h: V \to W$ is the inverse image of $\vec{0}_W$,

$$\mathrm{Ker}(h) = h^{-1}(\{\vec{0}_W\}) = \{x \in V \text{ such that } h(x) = \vec{0}_W\}.$$

The dimension of the nullspace is the map's *nullity*.

Note that $\mathrm{Ker}(f)$ and $\mathrm{Im}(f)$ are easily checked to be vector spaces. Warning: nullity cannot be translated in French by nullité (which does exist but means something else).

Theorem 25 *A linear map f is injective if and only if it has nullity 0, or equivalently,* $\ker(f) = \{\vec{0}\}$.

Proof 11 *Let $\vec{x}, \vec{y}$ be such that $f(\vec{x}) = f(\vec{y})$. Then $f(\vec{x}) - f(\vec{y}) = f(\vec{x} - \vec{y}) = \vec{0}$. Hence, $\vec{x} - \vec{y} \in \mathrm{Ker}(f)$. So, $\vec{x} - \vec{y} = \vec{0}$ and $\vec{x} = \vec{y}$.*

3.3.3 Computing images and kernels

If f is the linear map $f(x) = Mx$ then the image of f is the span of the columns of M.

The kernel of f can be computed by solving $Mx = 0$.

Theorem 26 *The dimension of the solution of an homogeneous system in echelon form is the number of its free variables.*

3.3.4 The rank-nullity theorem and its consequences

Theorem 27 (Rank-nullity theorem) *A linear map's rank plus its nullity equals the dimension of its domain. Rephrased: if $f: V \to W$ is linear then* $\dim(\mathrm{Ker}(f)) + \dim(\mathrm{Im}(f)) = \dim(V)$.

Proof 12 *Let $(e_1, \ldots, e_k)$ be a basis of* $\mathrm{Ker}(f)$. *Let us add vectors $f_1, \ldots, f_l$ so that $(e_1, \ldots, e_k, f_1, \ldots, f_l)$ is a basis of V (this can be done by Theorem 19). We claim that $(f(f_1), \ldots, f(f_l))$ is a basis of* $\mathrm{Im}(f)$.

Indeed, if $\lambda_1 f(f_1) + \cdots + \lambda_n f(f_n) = 0$ then $f(\lambda_1 f_1 + \cdots + \lambda_n f_n) = 0$. Hence, $\lambda_1 f_1 + \cdots + \lambda_n f_n \in \mathrm{Ker}(f)$. So, $\lambda_1 f_1 + \cdots + \lambda_n f_n = \mu_1 e_1 + \cdots + \mu_k e_k$ since $(e_1, \ldots, e_k)$ is a basis of $\mathrm{Ker}\, f$. *So, $\lambda_1, \ldots, \lambda_k = 0$. This proves that $f_1, \ldots, f_l$ are independent.*

If $y \in \mathrm{Im}(f)$ then $y = f(x)$ for some $x \in V$. Hence, $y = f(\mu_1 e_1 + \cdots + \mu_k e_k + \lambda_1 f_1 + \cdots + \lambda_n f_n) = \lambda_1 f(f_1) + \cdots + \lambda_n f(f_n)$. So, $f_1, \ldots, f_l$ span $\mathrm{Im}(f)$. *This proves our claim and the theorem.*

Note that the next theorem holds in particular for every endomorphism.

Theorem 28 *Let h be a linear map from a vector space V of dimension n to a vector space W of same dimension. Then the following properties are equivalent :*

1. *h has rank n;*
2. *h has nullity 0;*
3. *h is an injection;*
4. *h is a surjection;*
5. *h is a bijection;*
6. *For some basis $(\vec{b}_1, \ldots, \vec{b}_n)$ of V, $(h(\vec{b}_1), \ldots, h(\vec{b}_n))$ is a basis of V;*
7. *For every basis $(\vec{b}_1, \ldots, \vec{b}_n)$ of V, $(h(\vec{b}_1), \ldots, h(\vec{b}_n))$ is a basis of V.*

Proof 13 *By Theorem 27, (1) and (2) are equivalent. By Theorem 25, (2) and (3) are equivalent. Clearly, (1) and (4) are equivalent. Since (3) and (4) are equivalent, they are equivalent to (5). Clearly, (7) implies (6).*

Let us prove (6) implies (2). If $h(\vec{x}) = 0$, then $h(x) = h(\lambda_1\vec{b}_1 + \cdots + \lambda_n\vec{b}_n)$ $= \lambda_1 h(\vec{b}_1) + \cdots + \lambda_n h(\vec{b}_n)$. So, $\lambda_1 = 0, \ \ldots, \ \lambda_n = 0$ since $(h(\vec{b}_1), \ldots, h(\vec{b}_n)$ is a basis. Hence, $\vec{x} = 0$.

Let us prove (5) implies (7). If $(\vec{b}_1, \ldots, \vec{b}_n)$ is a basis, then let us prove that $f(\vec{b}_1), \ldots, f(\vec{b}_n)$ are independent. If $\lambda_1 f(\vec{b}_1) + \cdots + \lambda_n f(\vec{b}_n) = 0$ then $f(\lambda_1 b_1 + \cdots + \lambda_n \vec{b}_n) = \vec{0}$. Since f is a bijection, this implies $\lambda_1\vec{b}_1 + \cdots + \lambda_n\vec{b}_n = \vec{0}$. So, $\lambda_1, \ldots, \lambda_n = 0$. Let us prove that $(f(\vec{b}_1), \ldots, f(\vec{b}_n))$ spans V. Let $\vec{y}$ be in V. Then $\vec{y} = f(\vec{x})$ since f is a bijection. So, $f(\vec{x}) = \lambda_1 f(\vec{b}_1) + \cdots + \lambda_n f(\vec{b}_n)$ where the coefficients $\lambda_1, \ldots, \lambda_n$ are defined by $\vec{x} = \lambda_1\vec{b}_1 + \cdots + \lambda_n f\vec{b}_n$.

3.3.5 projection

Let us call *projection* any linear map p from V to V such that $p(p(x)) = p(x)$ for every $x \in V$.

Exercise 1 : A map p is a projection if and only if: $\mathrm{Ker}(f) \oplus \mathrm{Im}(f) = V$. Moreover, for every $\vec{x} \in V$, if $\vec{x} = \vec{y} + \vec{z}$ where $y \in \mathrm{Ker}(f)$ and $\vec{z} \in \mathrm{Im}(f)$ then $p(\vec{x}) = \vec{z}$.

3.3.6 Isomorphisms

Isomorphisms between spaces are maps that formalize the feeling that some spaces are "just like oneother".

An *isomorphism* between two vector spaces V and W is a linear map f from V to W that is also a bijection. Two spaces such that there exists an isomorphism between them are *isomorphic*.

Note that Theorem 28 applies in particular to isomorphism.

Theorem 29 *Two vector spaces over $\mathbb{R}$ of finite dimension are isomorphic if and only if they have same dimension. If V is a vector space over $\mathbb{R}$ of finite dimension, then it is isomorphic to $\mathbb{R}^n$ for some n.*

Proof 14 *If two spaces are isomorphic, then they clearly have same dimension. The proof is left to the reader as an exercise.*

The following theorem says that isomorphism preserve dimension.

Theorem 30 *Let ϕ be an isomorphism from a space V to a space W. Let V' be a subspace of V. Then V' and $\phi(V')$ have same dimension.*

3.3.7 Representing vectors and linear maps with matrices

We already know that to every matrix we can naturally associate a function. Here we prove the converse : up to the choice of a basis, every linear map can be represented by a matrix.

Let V be a vector space of dimension n and $B = (\vec{b}_1, \ldots, \vec{b}_n)$ be a basis. The *representation* of a vector $x \in V$ with respect to B is the vector of $\mathbb{R}^n$ whose components are the coefficients of the unique linear combination of the $\vec{b}_i$'s that equals x.

Suppose that V and W are vector spaces of dimensions n and m with bases B and D, and that $f: V \to W$ is a linear map. The *matrix that represents f with respect to B, D* is the $m \times n$ matrix whose column j is the representation of $f(\vec{b}_j)$ with respect to D.

Example : derivative of polynomials.

Theorem 31 *Let $f: V \to W$ be a linear map, and suppose that V is of dimension n with a basis B and W of dimension m with a basis D. If M is*

the representation of f with respect to B, D and X, Y are the representation of $x \in V, f(x) \in W$ with respect to B, D then:

$$Y = MX$$

Proof 15 *Let $\vec{x} = \lambda_1\vec{b_1} + \cdots + \lambda_n\vec{b_n}$ be a vector of V. Since f is linear we have $f(\vec{x}) = \lambda_1 f(\vec{b_1}) + \cdots + \lambda_n f(\vec{b_n}) = \lambda_1(f_{1,1}\vec{d_1} + \cdots + f_{m,1}\vec{d_m}) + \cdots + \lambda_n(f_{1,n}\vec{d_1} + \cdots + f_{m,n}\vec{d_m}) = (f_{1,1}\lambda_1 + \cdots + f_{1,n}\lambda_n)\vec{d_1} + \cdots + (f_{m,1}\lambda_1 + \cdots + f_{m,n}\lambda_n)\vec{d_m}$.*

Theorem 32 *Let $h, g: V \to W$ be linear maps represented with respect to bases B, D by the matrices M and N, and let t be a scalar. Then the map $h + g: V \to W$ is represented with respect to B, D by $M + N$, and the map $t \cdot h: V \to W$ is represented with respect to B, D by tM.*

Theorem 33 *A composition of linear maps is represented by the matrix product of the representatives.*

Theorem 34 *Let M be a non-singular matrix. Then there exists a unique matrix N such that $MN = NM = I$.*

Proof 16 *Since M is non-singular, by theorem 10 we know that $f: X \to MX$ is bijective. So, f must have a unique inverse map g. We claim that g is linear. Indeed, $g(\lambda x + \mu y)$ and $\lambda g(x) + \mu g(y)$ are equal, since their image by the bijection f is the same vector $\lambda x + \mu y$.*

Let N be the matrix of g with respect to the canonical basis of $\mathbb{R}^n$. Note that M is the matrix of f with respect to the same basis. Since $f \circ g = g \circ f = I$ and by Theorem 33, we have $MN = NM = I$.

The matrix N in the theorem above is the *inverse* of M, denoted by M^{-1}.

Practically, to compute M^{-1}, a method is to solve $Mx = y$ where x, y are general vectors. For instance, if $P = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix}$ this yields:

$$\begin{cases} 3x_1 + 2x_2 = y_1 \\ 2x_1 + \ \ x_2 = y_2 \end{cases}$$

$$\begin{cases} \ \ 3x_1 + x_2 = & y_1 \\ (1/2)x_1 \qquad = y_2 - (1/2)y_1 \end{cases}$$

$$\begin{cases} x_1 = -y_1 + 2y_2 \\ x_2 = \ 2y_1 - 3y_2 \end{cases}$$

Hence, $P^{-1} = \begin{pmatrix} -1 & 2 \\ 2 & -3 \end{pmatrix}$ is read from the solution.

Theorem 35 *If A, B are two non-singular $n \times n$ matrices then AB is non-singular and $(AB)^{-1} = B^{-1}A^{-1}$.*

3.3.8 Change of basis

Let $B = (\vec{b}_1, \ldots, \vec{b}_n)$ be a basis of a vector space, and $B' = (\vec{b}'_1, \ldots, \vec{b}'_n)$ be another basis of the same vector space. The *change of basis matrix from B to B'* is the $n \times n$ matrix P whose column j is b'_j represented with respect to B. It is convenient to think of B as the "old" basis, and B' as the "new" basis. So, P is obtained by expressing the new basis into the old one. Note that P is the matrix representing the identity with respect to B', B. It follows that P is invertible and that P^{-1} is the matrix representing the identity with respect to B, B'.

Theorem 36 *Suppose that B, B' are basis of a vector space and x is a vector of that space. If P is the change of basis matrix from B to B', if X, X' are the vectors representing x with respect to B, B' then $X = PX'$.*

For instance, if B is the standard basis of $\mathbb{R}^2$, if $B' = (\begin{pmatrix} 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix})$ then $P = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix}$. If $X = \begin{pmatrix} 5 \\ 4 \end{pmatrix}$ then, $X = PX'$, so $X' = P^{-1}X = \begin{pmatrix} -1 & 2 \\ 2 & -3 \end{pmatrix}\begin{pmatrix} 5 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$. This means that $\begin{pmatrix} 5 \\ 4 \end{pmatrix}$ has components $3, -2$ with respect to the basis B'. Indeed, $\begin{pmatrix} 5 \\ 4 \end{pmatrix} = 3 \cdot \begin{pmatrix} 3 \\ 2 \end{pmatrix} + -2 \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

Theorem 37 *Suppose that B, B' are basis of a vector space V, that D, D' are basis of a vector space W and that $f: V \to W$ is a linear map. Let P be the change of basis matrix from B to B' and Q be the change of basis matrix*

from D to D'. Let M, M' be the matrices that represent f with respect to B, D and B', D' respectively. Then:

$$M' = Q^{-1}MP$$

Proof 17 *Let $x \in V$ and let X, Y be the vectors representing $x, f(x)$ with respect to B, D respectively. Let X', Y' be the vectors representing $x, f(x)$ with respect to B', D'. Then by Theorems 36, 31, $Y = MX$, $Y' = M'X'$, $X = PX'$ and $Y = QY'$. Hence, $QM'X' = QY' = Y = MX = MPX'$. Hence, for every vector X' we have $QM'X' = MPX'$. This implies $QM' = MP$ and the result follows after multiplying by Q^{-1}.*

Example: To compute the matrix of the projection of $\mathbb{R}^2$ on the span of $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$ parallel to $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$, we can put $B' = \left\{ \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$. The matrix M' of p with respect to B', B' is $M' = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. The change of basis matrix from B to B' is $P = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix}$. Here, $Q = P$. Hence, $M' = P^{-1}MP$ and $M = PM'P^{-1} = \begin{pmatrix} -3 & 6 \\ -2 & 4 \end{pmatrix}$. We check $M^2 = M$, $M \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$ and $M \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 0$.

3.3.9 Rank of a matrix

The *rank* of an $m \times n$ matrix M is the rank of the linear map $f_M : X \to MX$.

Theorem 38 *Let M be an $m \times n$ matrix. Let A be an $m \times m$ matrix, and B be an $n \times n$ matrix. Suppose that A, B are non-singular. Then M, AM and MB have same rank.*

Proof 18 *AM has rank the dimension of $\mathrm{Im}(f_A \circ f_M)$. By Theorem , has same dimension than $\mathrm{Im}(f_M)$. So, M and AM have same rank. Similarly, M and MB have same rank.*

A consequence is that matrices obtained from one another by Gaussian row and Gaussian column operations have same rank because of Theorem13.

Theorem 39 *An $m \times n$ matrix has rank k if and only if there is a sequence of row and column Gaussian operations the transform M into : $\begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$.*

Proof 19 *By the Gaussian algorithm, any matrix M can be transformed into $\begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$. Since this matrix has rank k, it follows that M has rank k, and in particular that k is unique.*

Theorem 40 *The rank of an $m \times n$ matrix is equal to the dimension of the subspace of $\mathbb{R}^m$ spanned by its columns, and is equal to the dimension of the subspace of $\mathbb{R}^n$ spanned by its rows.*

Proof 20 *By the definition of the rank, it is clear that it equals the dimension of the column space. Indeed, the columns of a matrix are the image of the standard basis. By the preceding theorem, it follows that row and column operation are symmetric with respect to the rank. Hence, a matrix and its transpose must have same rank.*

Chapter 4

Determinants

Determinants are an algebraic construction that associates a real number to every square matrix, in such a way that the number equals 0 if and only if the matrix is singular.

4.1 The particular case of dimension 2

In dimension 2, it is "well known" that the determinant of M, denoted by $|M|$ is defined by

$$\left| \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right| = ad - bc.$$

Note that the mapping "det" can be viewed either as a mapping of M or a mapping of its columns.

Proposition 6 *for any columns vectors C_1, C_2 and C_3 of $\mathbb{R}^2$,*

- $\det(C_1, C_2) = -\det(C_2, C_1)$.
- $\det(C_1 + \alpha C_3, C_2) = \det(C_1, C_2) + \alpha \det(C_3, C_2)$.
- $\det(C_1, C_2 + \alpha C_3) = \det(C_1, C_2) + \alpha \det(C_1, C_3)$.

Consequently, if a matrix has two identical columns, then its determinant is 0. The properties stated above are summarized as "alternating multi-linearity".

We can study the effects of the Gaussian operations on columns:

- the determinant changes of sign if columns are exchanged ($C_1 \leftrightarrow C_2$).
- the determinant is unchanged ($C_k \to C_k + \alpha C_j$) with $j \neq k$).
- the determinant is multiplied by the same scalar ($C_k \to \lambda C_k$ with $\lambda \neq 0$).

Theorem 41 *A square matrix is non-singular if and only its determinant is non-singular, that is if and only if its determinant is not 0.*

Proposition 7 $\det(A) = \det(A^T)$.

Consequently, the results can be transposed for rows. In particular, the determinant is invariant under pivoting rows or columns.

Theorem 42 $\det(AB) = \det(A)\det(B)$.

Corollary 2 *Two similar matrices have same determinant.*

A consequence of Corollary2 and Theorem 52 is that for any matrix A and B that represent the linear mapping $f \in \mathcal{L}(E)$, $\det(A) = \det(B)$. This common number will be denoted by $\det(f)$.

4.2 The general case

The standard (and in general fastest) way to compute a determinant is by Gaussian elimination. But determinants can be computed by a other means, and are involved in several useful formulas, that compute the inverse of a non-singular matrix, or that solve linear systems for instance. Here it will be considered as the definition by induction of the determinant.

For any $n \times n$ matrix A, the $(n-1) \times (n-1)$ matrix formed by deleting row i and column j of A is the (i,j)-*minor* of A. The (i,j)-*cofactor* $t_{i,j}$ of T is $(-1)^{i+j}$ times the determinant of the (i,j)-minor of T.

Theorem 43 (Laplace Expansion of Determinants) *Where A is an $n\times n$ matrix, the determinant can be found by expanding by cofactors on row i or column j.*

$$\begin{aligned} |T| &= t_{i,1} \cdot a_{i,1} + t_{i,2} \cdot a_{i,2} + \cdots + t_{i,n} \cdot a_{i,n} \\ &= t_{1,j} \cdot a_{1,j} + t_{2,j} \cdot a_{2,j} + \cdots + t_{n,j} \cdot a_{n,j} \end{aligned}$$

We can initiate the process by considering that the determinant of $(1 \times 1)-$matrix is determined by $\det(a) = a$.

Remark 4 *If a matrix has a column of 0's, then its determinants is 0. The determinant of a triangular matrix is the product of the entries in its diagonal.*

It convenient to use the following notation: if $x_1, \ldots, x_n$ are vectors of $\mathbb{R}^n$, then we denote by $(x_1, \ldots, x_n)$ the matrix whose columns are $x_1, \ldots, x_n$.

Theorem 44 det *is an alternating n-linear map.*
If a matrix has two identical columns, then its determinant is 0.

The following theorem keeps track of the value of a determinant after Gaussian operations on columns of a matrix:

Theorem 45
- *the determinant changes of sign if columns are exchanged* $(C_i \leftrightarrow C_j$ *with* $j \neq i)$.
- *the determinant is unchanged* $(C_k \to C_k + \alpha C_j)$ *with* $j \neq k)$.
- *the determinant is multiplied by the same scalar* $(C_k \to \lambda C_k$ *with* $\lambda \neq 0)$.

Theorem 46 *A square matrix is non-singular if and only its determinant is non-singular, that is if and only if its determinant is not 0.*

Theorem 47 $\det(A) = \det(A^T)$.

A consequence of the theorem above is that the determinant is invariant under pivoting rows or columns.

Theorem 48 $\det(AB) = \det(A)\det(B)$.

Theorem 49 *Two similar matrices have same determinant.*

A consequence of the theorem above is that it is well defined to consider the determinant of an endomorphism by the determinant of one of its matrix representation.

Theorem 50 *A matrix M has rank r if and only if there exists an $r \times r$ submatrix of M that has a non-zero determinant and every $(r+1) \times (r+1)$ submatrix of M (if any) has determinant 0.*

4.2.1 Useful formulas involving determinants

Determinant par bloc. (to be written ...)

Definition 6 *The matrix* adjoint *to the square matrix T is*

$$\mathrm{adj}(A) = \begin{pmatrix} t_{1,1} & t_{2,1} & \dots & t_{n,1} \\ t_{1,2} & t_{2,2} & \dots & t_{n,2} \\ & \vdots & & \\ t_{1,n} & t_{2,n} & \dots & t_{n,n} \end{pmatrix} \tag{4.1}$$

where $T_{j,i}$ is the j,i cofactor.

Theorem 51 *When A is a square matrix, $A\,\mathrm{adj}(A) = \mathrm{adj}(A)A = \det(A)I$. In particular, if A is non-singular, $A^{-1} = |A|^{-1}\,\mathrm{adj}(A)$.*

4.2.2 Some determinants of interest

4.2.3 Similarity

The $n \times n$ matrices T and S are *similar* if there is a an $n \times n$ nonsingular matrix P such that $T = P^{-1}SP$.

Theorem 52 *Let A, B be two $n \times n$ square matrices. They are similar if and only if there exists a vector space V, two basis B, D of V and an endomorphism f of V such that A is the representation of f with respect to B and B is the representation of f with respect to D.*

Proof 21 *This is Theorem 37 rephrased.*

Proposition 8 *Let A, B be two $n \times n$ square matrices. If they are similar, then $\mathrm{Tr}(A) = \mathrm{Tr}(B)$.*

Chapter 5

diagonalization

5.1 Eigenvalues, eigenvectors

Definition 7 *Let V be a vector space and φ be an endomorphism of E. Let $\lambda \in \mathbb{R}$ be a scalar, λ is an* eigenvalue *of φ if there exists a nonzero vector $\vec{x} \in V$ such that $\varphi(\vec{x}) = \lambda\vec{x}$. Such a non-zero vector $\vec{x}$ is called an* eigenvector *associated to the corresponding eigenvalue.*

Note that an eigenvector is associated to a unique eigenvalue while an eigenvalue is associated to several eigenvectors.

Definition 8 *If λ is an eigenvalue, we will denote By V_λ the eigenspace associated to λ, defined by*

$$V_\lambda = \{\vec{x} \in V \text{ such that } \varphi(x) = \lambda\vec{x}\}.$$

Since we can remark that $V_\lambda = \mathrm{Ker}(\varphi - \lambda\,\mathrm{Id})$, it follows that the eigenspaces are vector subspace of V. Note that if λ is in $\mathbb{R}$, we can use similarly the same notation $V_\lambda = \{x \in V \text{ such that } \varphi(x) = \lambda x\}$. It is obvious that λ is an eigenvalue of φ if and only if $V_\lambda \neq \{\vec{0}\}$. Note also that, if $\vec{x}$ is an element of V_λ it is either the null vector or an eigenvector.

Remark 5 *Let V be a vector space and $\varphi \in \mathcal{L}(V)$. It is immediate that for any scalar λ, either*

- *$\lambda = 0$ and $V_0 = \mathrm{Ker}(\varphi)$ or*

- $\lambda \neq 0$ *and* $V_\lambda \subset \mathrm{Im}(\varphi)$.

Proposition 9 *If M is a triangular matrix, then the eigenvalues of M are the numbers on the diagonal of M.*

5.2 diagonalization

An important step towards diagonalization is the following proposition.

Proposition 10 *If $\lambda_1, \ldots, \lambda_k$ are k pairwise distinct scalars, then $V_{\lambda_1}, \ldots, V_{\lambda_k}$ are in direct sum, that is $V_{\lambda_1} + \cdots + V_{\lambda_k} = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_k}$.*

The proof is based on Vandermonde determinants in the general case but it is obvious in the case of two distinct eigenvalues. In view of Proposition 3, let us consider $\vec{x} \in V_{\lambda_1} \cap V_{\lambda_2}$, then we have $\varphi(\vec{x}) = \lambda_1\vec{x} = \lambda_2\vec{x}$. Consequently, $(\lambda_1 - \lambda_2)\vec{x} = \vec{0}$. Since the eigenvalues λ_1 and λ_2 are distinct, $\vec{x}$ is the null vector and the sum is direct.

Theorem 53 *Let h be an endomorphism of a space V of dimension n. The following conditions are equivalent:*

- *the sum the eigenspaces of h is V;*
- *the sum of the dimensions of the eigenspaces of h equals n;*
- *there exists a basis of V whose elements are eigenvectors of h;*
- *there exits a basis B of V such that the representation of h with respect to B is a diagonal matrix.*

Definition 9 *An endomorphism that satisfies the conditions in the theorem above is* diagonalizable.

We can define similarly the notions of *eigenvalues*, *eigenvectors* and *eigenspaces* for square matrices, since an $n \times n$ matrix can be viewed as an endomorphism of $\mathbb{R}^n$.

Remark 6 *A matrix is diagonalizable if and only if it is similar to a diagonal matrix.*

Remark 7 *Note that if the eigenvalues are knowns then Theorem 53 gives a process in order to check whether matrix is diagonalizable. It suffices to determine the eigenspaces and their dimension.*

Example, let us consider

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

For any real number λ, we recall that λ is an eigenvalue if and only if the following linear system admits a non-trivial solution:

$$\begin{cases} 0x_1 & +x_2 & = \lambda x_1 \\ 0x_1 & +0x_2 & = \lambda x_2 \end{cases} \Leftrightarrow \begin{cases} x_2 & = \lambda x_1 \\ 0 & = \lambda x_2 \end{cases}$$

It is immediate that $E_\lambda = \{\vec{0}\}$ if $\lambda \neq 0$ and $E_\lambda = \text{span}(1, 0)$ (dimension 1). We can therefore conclude that M is not diagonalizable.

If there exists some condition on the iterates of a linear mapping, we can deduce a necessary condition on the eigenvalues. For example, if $f^2 = Id$, we can translate this condition as $\lambda^2 = 1$ for any eigenvalue of f.

The next section will give us a systematic method in order to look for the eigenvalues.

5.3 Characteristic polynomial

The *characteristic polynomial* of a square matrix A is the determinant $\chi_A(X)$ of the matrix $A - XI$, where X is a variable. The *characteristic equation* is $|A - xI| = \chi_A(x) = 0$.

The characteristic polynomial of an endomorphism f is the characteristic polynomial of the representation of f with respect to any basis $\mathcal{B}$ (the result does not depend on the choice of $\mathcal{B}$).

Proposition 11 *A scalar λ is an eigenvalue of a matrix (respectively of an endomorphism) if and only if it is a root of its characteristic polynomial.*

Since it is easy to compute the determinant of a triangular matrix, we can write an alternative proof of the triangular case.

Corollary 3 *If M is a triangular matrix, then the eigenvalues of M are the numbers on the diagonal of M.*

Proposition 12 *Let M be a square matrix. Then for any λ, $\dim V_\lambda \leq mult(\lambda)$ where $mult(\lambda)$ is the multiplicity of λ in the characteristic polynomial.*

For example, if $\chi_A(x) = (x-7)^3(x+5)^2$, it means that there are two eigenvalues 7 of multiplicity 3, and -5 of multiplicity 2. So we know that $\dim V_7 \leq 3$ and $\dim V_{-5} \leq 2$ but also, that 2 is not an eigenvalue, which implies that $\dim V_2 = 0$.

Theorem 54 *Let $M = (m_{i,j})$ be a square matrix of size n. Then χ_M is a polynomial of degree n which can be written as*

$$\chi_M = (-1)^n \lambda_n + a_{n_1}\lambda^{n-1} + \ldots + a_0$$

where $a_0 = \det(M)$ and $a_{n-1} = (-1)^{n-1}\operatorname{Tr}(M)$.

For example, if $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ It follows that $\chi_M(\lambda) = \lambda^2 - (a+d)\lambda + (ad - bc)$.

Corollary 4 *An $n \times n$ matrix (respectively an endomorphism of a vector space of dimension n) has at most n distinct eigenvalues.*

Warning: the converse of the theorem above is false: an $n \times n$ matrix may be diagonalizable while having less than n pairwise distinct eigenvalues.

Theorem 55 *Let $M = (m_{i,j})$ be a square matrix of size n (respectively $\varphi \in \mathcal{L}(E)$ where E is a vector space of dimension n). We denote by $\lambda_1, \ldots, \lambda_k$ the roots of χ_M (respectively χ_φ). We denote by n_i the multiplicity of λ_i in the polynomial. Either*

- *$n_1 + \ldots + n_k < n$. Then M (respectively φ) is not diagonalizable.*

- $n_1 + \ldots + n_k = n$ *and there exists some* i *such that* $\dim(E_{\lambda_i}) < n_i$. *Then* M *(respectively* φ*) is not diagonalizable.*
- $n_1 + \ldots + n_k = n$ *and for all* i, $\dim(E_{\lambda_i}) = n_i$. *Then* M *(respectively* φ*) is diagonalizable. In addition, we have* $V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_k}$ *and a basis of diagonalization can be build by merging the basis of* $V_{\lambda_1}, \ldots, V_{\lambda_k}$.

In particular, if the decomposition of the characteristic polynomial includes a an irreducible polynomial of degree 2, we can conclude that the diagonalization is not possible. The following result is used frequently in the exercises.

Example, if we consider $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. It is easy to check that $\chi_A(X) = X^2 + 1$. But this polynomial has no real root, so it can not be decomposed. It is irreducible, therefore A is not diagonalizable.

Theorem 56 *If an* $n \times n$ *matrix has* n *pairwise distinct eigenvalue then it is diagonalizable.*

Proof 22 *Let us first remark that for any eigenvalue* $\lambda_1, \ldots, \lambda_n$, *the multiplicity equals* 1 *and consequently* $1 \leq \dim V_{\lambda_i} \leq 1$, *which implies that* $\dim(V_{\lambda_i}) = 1$. *So the sum of the dimension of the eigenspaces equals the dimension of the space.*

Theorem 57 (Cayley-Hamilton) *Let* M *be a square matrix. Then* $\chi_M(M) = 0$.

Let us first present an example in order to understand the conclusion. Let us consider the matrix $\begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}$. It is easy to compute the characteristic polynomial:

$$\begin{vmatrix} 2-\lambda & 1 \\ 3 & 4-1\lambda \end{vmatrix} = (2-\lambda)(5-\lambda) - 1 \times 3 = 5 - 6\lambda + \lambda^2$$

With the convention, that $M^0 = \mathrm{Id}$, $\chi_M(M) = M^2 - 6M + 5\,\mathrm{Id}$. Therefore,

$$\chi_M(M) = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}^2 - 6\begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix} + 5\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 7 & 6 \\ 18 & 19 \end{pmatrix} + \begin{pmatrix} -12 & -6 \\ -18 & -24 \end{pmatrix} + \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Note that the result is obvious if the matrix is diagonal or even diagonalizable.

Appendix A

Algebraic prerequisite

A.1 Polynomials

Let us introduce the algebraic concept of polynomial which is closely related to the concept of polynomial functions.

Let us call *polynomial* any sequence $P = (a_n)_{n\in\mathbb{N}}$ of real numbers such that there exists an integer d satisfying $a_i = 0$ for every $i > d$. The numbers a_i are the *coefficients* of P.

The polynomial $(0, 1, 0, \ldots)$ is often[1] denoted by X, leading to a set of polynomial denoted by $\mathbb{R}[X]$. Note that X is not any more a real number, it is an algebraic object.

In this case, $(0, 0, 1, 0, \ldots)$ is denoted by X^2, $(0, 0, 0, 1, 0, \ldots)$ is denoted by X^3, and more generally $X^k = (0, \ldots, 0, 1, 0 \ldots) = (\delta(n, k))_{n\in\mathbb{N}}$ where $\delta(n, k)$ is the *Kronecker's delta*, equal to 1 if $n = k$ and equal to 0 otherwise.

Those notations allow to establish a link between the formal definition of the polynomial $(1, -1, 0, 1, 0, 0 \ldots)$ and a "concrete" polynomial. The sequence $(1, 2, -1, 0, 1, 0, 0 \ldots)$ corresponds to the polynomial $1+2X-X^2+X^4$. Conversely, the sequence is the result of "coding": $1 - X + X^3$ will be coded by the sequence $(1, -1, 0, 1, 0, 0 \ldots)$. In view of this "idenfication", we will say that $1 - X + X^3$ is a polynomial.

The largest integer d such that $a_d \neq 0$ is the *degree* of P, denoted by $\deg(P)$. If every coefficient of P is zero, then P is the zero polynomial, denoted by 0. Its degree is defined to be $-\infty$. The polynomial of degree

[1] alternatively, if we denote $(0, 1, 0, \ldots)$ by Y, the set of polynomial will be $\mathbb{R}[Y]$.

0 whose first coefficient is $a \in \mathbb{R}$ and every other coefficient is 0 is simply denoted by a. Such a polynomial is said to be *constant*. It follows easily that every non-zero polynomial can be written $a_0 + a_1X + \ldots + a_dX^d$ where d is its degree.

Let us define formal concepts of sum and products on those algebraic object. It is important to notice that they coincide with the usual operations on polynomial functions.

The *sum* of two polynomials $P = (a_n)_{n\in\mathbb{N}}$ and $Q = (b_n)_{n\in\mathbb{N}}$ is $P + Q = (a_n + b_n)_{n\in\mathbb{N}}$, which is in fact the usual sum. It is easy to see that $P + Q$ is a polynomial and that[2] that $\deg(P + Q) \leq \max(\deg(P), \deg(Q))$.

The *product* of a scalar λ by a polynomial $P = (a_n)_{n\in\mathbb{N}}$ is $\lambda P = (\lambda a_i)_{n\in\mathbb{N}}$. It is easy to see that λP is a polynomial and that $\deg(\lambda P) = \deg(P)$ if $\lambda \neq 0$ while $\deg(\lambda P) = -\infty$ if $\lambda = 0$. Note that $X^2 = X \times X$.

The *product* of two polynomials $P = (a_n)_{n\in\mathbb{N}}$ and $Q = (b_n)_{n\in\mathbb{N}}$ is $PQ = (\sum_{i+j=n} a_ib_j)_{n\in\mathbb{N}}$. It is easy to see that PQ is a polynomial and that $\deg(PQ) = deg(P) + deg(Q)$.

Let us check that those rules lead to a vector space structure:

Theorem 58 *For every polynomials P, Q, R and every scalars λ, μ we have:*

- $P + (Q + R) = (P + Q) + R$, $P + Q = Q + P$, $P + 0 = P$, $P - P = 0$;
- $P \cdot (QR) = (PQ) \cdot R$, $PQ = QP$, $1P = P$;
- $\lambda(P+Q) = \lambda P+\lambda Q$, $(\lambda\mu)P = \lambda(\mu P)$, $(\lambda+\mu)P = \lambda P+\mu P$, $\lambda(PQ) = (\lambda P) \cdot Q = P \cdot (\lambda Q)$;
- $P \cdot (Q + R) = PQ + PR$.

Note that 1 might denote the real number 1, or the polynomial $(1, 0, 0, \ldots)$. But in both cases, $1P = P$.

Theorem 59 *If $PQ = 0$ then either $P = 0$ or $Q = 0$. Also, if $PA = PB$ for some P non equal to 0, then $A = B$.*

Proof 23 : degree.

[2]Note that the formula is still true if one of the polynomial values zero in view of the convention introduced for the zero polynomial

As usually, we can compose the polynomials P and Q. If the polynomial $P = a_0 + a_1X + \ldots + a_dX^d$, the composition $P \circ Q$ will be defined as $a_0 + a_1Q + \ldots + a_dQ^d$. It can be denoted by $P \circ Q = P(Q)$. In particular, we can check $P(X) = P$ and for any real number $P(a)$ is constant.

A polynomial P is *invertible* if there exists a polynomial Q such that $PQ = 1$. The following theorem shows that except for some very special cases, a polynomial is not invertible.

If A is equal to PB, then A is a *multiple* of B and B is a *divisor* of A.

Proposition 13 *Let P be an invertible polynomial. Then $P = a$ where a is a non-zero real number.*

A.2 Euclidean division

Let us define the concept of *Euclidean division* of A by B, which is based on the following theorem.

Theorem 60 *Let A and $B \neq 0$ be two polynomials. There exists a unique Q and a unique R such that $A = BQ + R$ and $\deg(R) < \deg(B)$.*

For example, we can write $X^2 + X - 1 = X(X+1) - 1$. How to determine Q and R ? By solving linear systems : For example, if we consider $A = X^5 + 1$ and $B = X^2 + 1$, let us consider the decomposition $A = BQ + R$ or $A - R = BQ$. Since $\deg(B) \leq 1$, we know that $\deg(A - R) = 5$, therefore $\deg(BQ) = 5$, and consequently Q has degree 3. In view of the degrees, we can denote $B = aX^3 + bx^2 + cX + d$ and $R = eX + f$.

$$A = BQ + R \text{ means } X^5 + 1 = (X^2 + 1)(aX^3 + bx^2 + cX + d) + eX + f.$$

In particular $(X^2 + 1)(aX^3 + bx^2 + cX + d) + eX + f$ can be decomposed as aX^5+ terms of degree less than 5. We can deduce that $a = 1$. We need to solve

$$X^5 + 1 = (X^2 + 1)(X^3 + bx^2 + cX + d) + eX + f.$$

It is easy to compute that the right-hand side term is equal to

$$(X^5 + bx^4 + cX^3 + dX^2) + (X^3 + bx^2 + cX + d) + (eX + f)$$

which can grouped as $X^5+bx^4+(c+1)X^3+(d+b)X^2+(c+e)X+(d+f)$. So $A = BQ+R$ means that $X^5+1 = X^5+bx^4+(c+1)X^3+(d+b)X^2+(c+e)X+f$, leading to $b = 0$, $c + 1 = 0$, $d + b = 0$, $c + e = 0$ and $d + f = 0$. This linear system is easy to solve, $b = 0$, $c = -1$, $d = 0$, $e = 1$ and $f = 1$.

$$X^5 + 1 = (X^2 + 1)(X^3 - X) + (X + 1)$$

The pair (Q, R) in the above theorem is the result of the *Euclidean division* of A by B. The polynomial Q is the *quotient* and R is the *rest*. If the rest in division of A by B is 0, then B is a *divisor* of A, B *divides* A, and A is a *multiple* of Q.

Remark 8 *Let A, B be two polynomials. One has the equivalence*

- *A is a divisor of B and B is a divisor of A*
- *there exists $k \in \mathbb{R}^*$ such that $A = kB$.*

A.3 Arithmetic of polynomials

If A, B are such that $A = kB$ where $k \in \mathbb{R}^*$, then they *equal up to the multiplication by a constant*. A *unitary* polynomial is a polynomial whose coefficient of highest degree is 1. It is easily checked that every non-zero polynomial is equal to a unitary polynomial, up to the multiplication by a constant.

Let us denote by $D(P)$ the sets of all the divisors of P. In particular P is a divisor of P. In order to understand this set, let us remark that for any P, both $1 \in D(P)$ and $P \in D(P)$ while 0 consider some example

- if $P = 0$ then $D(P) = \mathbb{R}[X]$.
- if $P = 1$ then $D(P) = \mathbb{R}[X] \setminus \{0\}$.
- if $P = X$ then $D(P) = \mathbb{R}[X] \setminus \{\text{constant polynomials}\}$.

Note that $D(P) = D(Q)$ if and only if they are equal up to the multiplication by a constant (cf. Remark 8).

An important problem is to have a description of $D(A) \cap D(B)$ when A, B are polynomial. This set is the set of the common divisors of A and B,

which is a step in order to define the concept of Greatest Common Divisor. The following theorem shows that the computation of $D(A) \cap D(B)$ can be replaced by a simpler computation:

Theorem 61 *Let A, B be polynomials such that $A = BQ + R$. Then*

$$D(A) \cap D(B) = D(R) \cap D(B).$$

By the above theorems, the set $D(A) \cap D(B)$ can be computed as follows: if $B = 0$, then it is $D(A)$. Else, we compute the quotient and the rest of the euclidean division of A by B, and we go on with the simpler problem $D(B) \cap D(R)$. At the end of this process, we will find a polynomial C such that $D(A) \cap D(B) = D(C)$, that is such that every common divisor of A, B is a divisor of C.

Note that

- if both A and B are equal to zero then $C = 0$ and C is unique;
- if either A or B are different from zero, as noticed previously this polynomial C is unique "up to the multiplication by a constant". In order to have unicity, we require in addition C to be unitary.

In any case, this unique polynomial C is the *GCD of A, B* (GCD means *Greatest Common Divisor*) denoted by $\mathrm{GCD}(A, B)$.

Theorem 62 (Bézout[3]) *Let A, B be two polynomials where $B \neq 0$. Then there exist polynomials U, V such that $UA + VB = \mathrm{GCD}(A, B)$.*

Polynomial such that $\mathrm{GCD}(A, B) = 1$ are said to be *relatively prime*.

Remark 9 *Note that, for any $a \in \mathbb{R}$, for any polynomial P, $(X - a)$ is a divisor of P if and only if $P(a) = 0$.*

A polynomial P (of degree n) is said to be *entirely decomposed* on $\mathbb{R}[X]$ if there exits some $(\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ such that

$$P = a_n(X - \alpha_1) \ldots (X - \alpha_n).$$

Remark 10 *Note that, for any polynomials P and Q such that P and Q are entirely decomposed on $\mathbb{R}[X]$. Then P and Q are relatively prime if and only if they have no common root (cf. Page 47). Such a property can be extended in the general case if we allow complex root.*

The Bézout's Theorem for relatively prime polynomials is so important that it is worth stating separately:

Theorem 63 (Bézout) *Two polynomials are relatively primes if and only if there exist polynomials U, V such that $UA + VB = 1$.*

The following theorem might seems trivial, but it relies on all the construction above. It says that in some respect, polynomials behave like integers.

Theorem 64 (Gauss) *Let us consider A, B two relatively prime polynomials.*

- *If A divides BC then A divides C.*
- *If A, and B are both divisors of C, then AB is a divisor of C.*
- *If in addition A is relatively prime to C, then A is relatively prime to BC.*

Proof 24 : Let U, V be such that $AU + BV = 1$, leading to $C = AUC + BVC$.

- Since A divides BC, there exists P such that $BC = AP$. So, $C = AUC + VAP = A(UC + VP)$. Hence, A divides C.
- Also, $C = AP_1$ and $C = BP_2$. So, $C = AUC + BVC = AU(BP_2) + BV(AP_1) = AB(UP_2 + VP_1)$. So, AB is a divisor of C.
- In addition, there exits $\tilde{U}, \tilde{V}$ such that: $A\tilde{U} + C\tilde{V} = 1$. We can compute

$$(AU + BV)(A\tilde{U} + C\tilde{V}) = 1^2 = 1$$

Let us develop $A^2U\tilde{U} + AUC\tilde{V} + BVA\tilde{U} + BVC\tilde{V} = 1$. So,

$$A(AU\tilde{U} + UC\tilde{V} + BV\tilde{U}) + BC(C\tilde{V}) = 1,$$

so A and BC are relatively prime.

Every non-zero polynomial P have divisors: itself (but also the polynomials equal to P up to the multiplication by non-zero scalar), and the non-zero scalars. A non-constant polynomial is *irreducible* if it has no other divisor.

So, a non-constant polynomial P is irreducible if and only there exists a divisor Q such that $0 < \deg(Q) < \deg(P)$.

It is easily seen that two unitary irreducible polynomials are relatively prime.

Theorem 65 *Let us consider P a polynomial, $P \in \mathbb{R}[X]$ of degree $n \geq 1$, we can discuss the irreducibility of P :*

- *if $n \geq 2$, P is not irreducible;*
- *if $n = 2$, P is irreducible if and only if it has no real roots;*
- *if $n = 1$, P is irreducible.*

For example, if $P = X^2+1$ was not irreducible, there would exist a divisor Q such that $0 < \deg(Q) < 2$. So Q can be written $X - a$, and in particular $P(a) = Q(a) = 0$ which is impossible.

Theorem 66 *Let A be a polynomial. Then A is equal to a product of irreductible polynomial. This decomposition is unique up to a permutation of its terms and up to a multiplication of its terms by non-zero scalars.*

A.4 Roots of a polynomial

If P is a polynomial, there is a function naturally associated to P:

A *root* of a polynomial is a real x such that $P(x) = 0$.

Theorem 67 *Let P be a polynomial. Then a is a root of P if and only if $(X - a)$ divides P.*

Proof 25 : By the Euclidean division: $P = (X - a)Q + R$ where $\deg(R) < deg(X - a)$. Hence, R is a number, so $P(a) = R$ which implies the result.

The *multiplicity* of a root of polynomial P is the greatest k such that $(X - a)^k$ divides P. If $P = \sum_{k=0}^{k=n} a_k X^k$, then we denote by P' the *derivative* of P, defined by $P' = \sum_{k=1}^{k=n} k a_k X^{k-1}$. By $P^{(k)}$ we denote the polynomial obtained from P after k derivatives.

Theorem 68 (Taylor) *Let P be a polynomial of degree n and a be a real number. Then:*

$$P(X) = \sum_{k=0}^{n} \frac{P^{(k)}(a)}{k!}(X-a)^k$$

Theorem 69 *Let a be the root of a polynomial P. Then the multiplicity of a is k if and only if $P(a) = \cdots = P^{(k-1)}(a) = 0$ and $P^{(k)}(a) \neq 0$.*

Index