Linear Algebra

Math -UA 9140

Lecture Notes

Joachim Lebovits

Summer 2021

Contents

Part I

Matrices Calculus

Chapter 1

Matrices Calculus

In this course $\mathbf{K}$ denotes either the set of real numbers, denoted $\mathbf{R}$ or the set of real numbers, denoted $\mathbf{C}$.

1.1 Definition

Definition 1.1.1 (Matrix). *A $p \times q$ matrix with coefficients in $\mathbf{K}$ is a rectangular array of numbers of $\mathbf{K}$ with p rows and q columns. The coefficients of the matrix are also called terms or elements of the matrix.*

We denote $\mathscr{M}(p, q, \mathbf{K})$ or $\mathscr{M}_{p,q}(\mathbf{K})$ the set of all $p \times q$ matrices with coefficients in $\mathbf{K}$. In the case where $p = q$, we use the shorthand notation $\mathscr{M}_p(\mathbf{K})$.

1.2 Notations

In order to represent a $p \times q$ matrix A, we usually use the following notation

$$A := \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,q} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,q} \\ \vdots & \vdots & & \vdots \\ a_{p,1} & a_{p,2} & \cdots & a_{p,q} \end{pmatrix}.$$

Remark 1.2.1. *it is important to respec the convention on the order of indices: we denote $a_{i,j}$ the coefficient of the matrix A which is on the i^{th} row and j^{th} column.*

When one wants to give the coefficients of A with a formula, one can also use a condensed notation:

$$A := (a_{i,j})_{1 \leqslant i \leqslant p, 1 \leqslant j \leqslant q} \qquad \text{or} \qquad A := (a_{i,j})_{(i,j) \in [\![1,p]\!] \times [\![1,q]\!]}.$$

where $[\![1, p]\!]$ stands for $\{1, 2, \cdots, p\}$ for any p in $\mathbf{N}^*$.

Example 1.2.2. *Denote $A := (i - j)_{1 \leqslant i \leqslant 2, 1 \leqslant j \leqslant 3}$. It is clear that:*

$$(i - j)_{1 \leqslant i \leqslant 2, 1 \leqslant j \leqslant 3} = \begin{pmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \end{pmatrix}.$$

When there is no ambiguity because we do know the “dimensions” p and q of the matrix A we simply write $A := (a_{i,j})$. In particular, we will denote (0), and one will call *null matrix*, the matrix fir which all coefficients equal 0. Hence there is a null matrix for every couple of integers (p, q). They are all denoted the same way. It does not provoke any confusion unless we do not take care to the context.

1.3 Particular Matrices

When $p = 1$, one says that A is row-matrix. In this case we write:

$$A = (a_1, a_2, \cdots, a_q).$$

When $q = 1$, one says that A is column-matrix. In this case we write:

$$A := \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}.$$

When $p = q$, one says that A is a *square matrix* of order p.
Let $A := (a_{i,j})$ be a square matrix of order p. The *diagonal* of the matrix A is the line constituted with the p coefficients $a_{1,1}, a_{2,2}, \cdots a_{p,p}$. The coefficients on the diagonal are also walled *diagonal coefficients* of A.
A diagonal matrix is a square matrix for which all the non-diagonal terms equal 0. in other words $a_{i,j} = 0$ for all couple (i, j) such that $i \neq j$. This matrix can therefore be written under the form:

$$A := \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & a_p \end{pmatrix}.$$

One also writes $A = \mathrm{diag}(a_1, a_2, \cdots a_p)$.
A upper triangular (*resp.* lower triangular) matrix is a square matrix $A := (a_{i,j})$ for which all the terms that are below (*resp.* above) the diagonal equal 0. In other words $a_{i,j} = 0$ for all $i > j$ (*resp.* for all $i < j$). thus it has the following form:

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,q} \\ 0 & a_{2,2} & \cdots & \vdots \\ \vdots & & \ddots & a_{p-1,p} \\ 0 & \cdots & 0 & a_{p,p} \end{pmatrix}, \qquad (\textit{resp.} \qquad \begin{pmatrix} a_{1,1} & 0 & \cdots & 0 \\ a_{2,1} & a_{2,2} & \cdots & \vdots \\ \vdots & & \ddots & 0 \\ a_{p,1} & \cdots & a_{p-1,p} & a_{p,p} \end{pmatrix}).$$

1.4 Operations on Matrices

We will say that two matrices $A := (a_{i,j})_{(i,j)\in[\![1,p]\!]\times[\![1,q]\!]}$ and $B := (b_{i,j})_{(i,j)\in[\![1,p']\!]\times[\![1,q']\!]}$ are equal if and only if they have the same "dimensions" p and q and the same coefficients. In other words, if:

$$\forall (i, j) \in \{1, \cdots, p\} \times \{1, \cdots, q\},\ a_{i,j} = b_{i,j}.$$

First Operation: Transpose of a Matrix

Definition 1.4.1 (Transpose of a matrix). *Let $A := (a_{i,j})_{(i,j)\in[\![1,p]\!]\times[\![1,q]\!]}$ be a matrix with coefficients in $\mathbf{K}$.*
The transpose matrix of A, denoted ${}^tA$, is the matrix of $\mathscr{M}_{q,p}(\mathbf{K})$ defined by the following formula:

$$({}^tA)_{i,j} := a_{j,i}.$$

Example 1.4.2. *The transpose of the matrix* $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ *is the matrix* $\begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$.

We easily see that ${}^t({}^tA) = A$.

Second Operation: Multiplication of a Matrix by a Scalar

Definition 1.4.3 (Product of a Matrix by a Scalar). *The product of a matrix* $A := (a_{i,j})_{(i,j)\in[\![1,p]\!]\times[\![1,q]\!]}$ *by a scalar* λ *in* $\mathbf{K}$ *is the matrix of* $\mathscr{M}_{p,q}(\mathbf{K})$*, denoted* $\lambda \cdot A$*, which is defined by the following formula:*

$$(\lambda \cdot A)_{i,j} := \lambda \cdot a_{i,j}.$$

We have ${}^t(\lambda \cdot A) = \lambda \cdot {}^tA$. Morover, one can have $\lambda \cdot A = (0)$ if and only if $\lambda = 0$ or if $A = (0)$.

Third Operation: Addition of two Matrices

Definition 1.4.4 (Addition of two Matrices). *The sum of two matrices* $A := (a_{i,j})_{(i,j)\in[\![1,p]\!]\times[\![1,q]\!]}$ *and* $B := (b_{i,j})_{(i,j)\in[\![1,p]\!]\times[\![1,q]\!]}$ *of* $\mathscr{M}_{p,q}(\mathbf{K})$ *is the matrix of* $\mathscr{M}_{p,q}(\mathbf{K})$*, denoted* $A + B$*, and defined by the following formula:*

$$(A + B)_{i,j} = a_{i,j} + b_{i,j}.$$

Example 1.4.5. *We hence have:*

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 1+7 & 2+8 & 3+9 \\ 4+10 & 5+11 & 6+12 \end{pmatrix} = \begin{pmatrix} 8 & 10 & 12 \\ 14 & 16 & 18 \end{pmatrix}.$$

It is easy to see that the following rules hold.

$$(\lambda + \mu) \cdot A = \lambda \cdot A + \mu \cdot A, \qquad \lambda \cdot (A + B) = \lambda \cdot A + \lambda \cdot B, \qquad {}^t(A + B) = {}^tA + {}^tB. \tag{1}$$

The readers who are familiar with Group Theory noticed that $(\mathscr{M}_{p,q}(\mathbf{K}), +)$ is an abelian group.

Fourth Operation: Multiplication of two Matrices

Definition 1.4.6 (Multiplication of two Matrices). *The product of a matrix* $A := (a_{i,j})_{(i,j)\in[\![1,p]\!]\times[\![1,q]\!]}$ *by a matrix* $B := (b_{k,l})_{(k,l)\in[\![1,q]\!]\times[\![1,r]\!]}$ *is the matrix* $C := (c_{i,l})_{(i,l)\in[\![1,p]\!]\times[\![1,r]\!]}$ *defined by the following formula:*

$$c_{i,l} = \sum_{k=1}^{q} a_{i,k} b_{k,l}. \tag{2}$$

Example 1.4.7. *We hence have:*

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \cdot \begin{pmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{pmatrix} = \begin{pmatrix} 1\cdot 7 + 2\cdot 9 + 3\cdot 11 & 1\cdot 8 + 2\cdot 10 + 3\cdot 12 \\ 4\cdot 7 + 5\cdot 9 + 6\cdot 11 & 4\cdot 8 + 5\cdot 10 + 6\cdot 12 \end{pmatrix}$$

$$= \begin{pmatrix} 7 + 18 + 33 & 8 + 20 + 36 \\ 28 + 45 + 66 & 32 + 50 + 72 \end{pmatrix} = \begin{pmatrix} 58 & 64 \\ 139 & 154 \end{pmatrix}.$$

Remark 1.4.8. *In order to define the product of a matrix A by a matrix B, we need that the number of columns of A equals the number of rows of B. The matrix AB has as much rows as A and as much columns as B.*

Example 1.4.9. *Let x_1, x_2, y_1 and y_2 be real numbers. Let's compute the two following products.*

$$(x_1, x_2) \cdot \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = (x_1 y_1 + x_2 y_2), \qquad \textit{and} \qquad \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \cdot (x_1, x_2) = \begin{pmatrix} y_1 x_1 & y_1 x_2 \\ y_2 x_1 & y_2 x_2 \end{pmatrix}.$$

It is easy to see that the following rule hold.

$$\lambda \cdot (AB) = (\lambda \cdot A) \cdot B = A \cdot (\lambda \cdot B). \tag{3}$$

Proposition 1.4.10. *The product of two matrices is an associative law. More precisely, if A is a matrix in $\mathscr{M}_{p,q}(\mathbf{K})$, B is a matrix in $\mathscr{M}_{q,r}(\mathbf{K})$ and C is a matrix in $\mathscr{M}_{r,s}(\mathbf{K})$ then:*

$$(AB)\,C = A\,(BC).$$

Proof. To be filled! □

Proposition 1.4.11. *For any matrices A in $\mathscr{M}_{p,q}(\mathbf{K})$ and B in $\mathscr{M}_{q,r}(\mathbf{K})$, we have:*

$${}^{t}(AB) = {}^{t}B\,{}^{t}A.$$

Proof. To be filled! See [CPY96, p.16 Proposition I.2] □

Proposition 1.4.12. *The product of matrices is distributive over addition. More precisely, if A and B are matrices of $\mathscr{M}_{p,q}(\mathbf{K})$ and C and D are matrices of $\mathscr{M}_{q,r}(\mathbf{K})$, then*

$$A \cdot (C + D) = A \cdot C + A \cdot D \qquad \textit{and} \qquad (A + B) \cdot C = A \cdot C + B \cdot C.$$

Proof. To be filled! See [CPY96, p.16 Proposition I.2]
To be filled! See [CPY96, p.16 Proposition I.2]
To be filled! See [CPY96, p.16 Proposition I.2] □

1.5 The ring of Square Matrices

From the previous propositions we deduce that the set $(\mathscr{M}_p(\mathbf{K}), +, \cdot)$ of square matrices of order p, endowed with the addition and the multiplication of matrices is a ring [1]. Its multiplicative identity is

[1] A ring is a set $\mathscr{R}$ equipped with two binary operations, denoted $+$ and $\cdot$, satisfying the following three sets of axioms, called the ring axioms:

1. $(\mathscr{R}, +)$ is an abelian group. In other words:

    (i) $(a + b) + c = a + (b + c)$, for all a, b, c in $\mathscr{R}$ (that is, $+$ is associative).

the identity matrix of order p, denoted I_p (also denoted I when there is no risk of confusion about the value of p), and defined by setting:

$$I_p := (\delta_{i,j})_{1\leqslant i,j\leqslant p} := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix},$$

where $\delta_{i,j}$ denotes the Kronecker symbol which value is 1 if $i = j$ and 0 otherwise. One can easily check that:

$$I_p \cdot A = A \cdot I_p = A,$$

for every $p \times p$ matrix A.

Remark 1.5.1. *1. Note that $(\mathscr{M}_p(\mathbf{K}), \cdot)$ is not commutative if $p > 1$. Hence we have:*

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 2 \end{pmatrix}.$$

2. Moreover $(\mathscr{M}_p(\mathbf{K})+, \cdot)$ is not an integral domain. That means that one can find two non zero matrices such that their product equals the null matrix. Hence:

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0_{\mathscr{M}_2(\mathbf{K})}.$$

1.6 Invertible Matrices

Definition 1.6.1 (Invertible Matrix). *A matrix A of $\mathscr{M}_p(\mathbf{K})$ is said to be invertible, if there exists a matrix B in $\mathscr{M}_p(\mathbf{K})$ such that:*

$$B \cdot A = A \cdot B = I_p.$$

The matrix B is denoted A^{-1} and is called invert of the matrix A. We also say in this case that A is a regular matrix.

Remark 1.6.2. *A matrix can be invertible only if it is a square matrix.*

Denote $\mathbf{GL}_p(\mathbf{K})$ (for general linear group of degree p) the set of all invertible matrices of $\mathscr{M}_p(\mathbf{K})$.

(*ii*) $a + b = b + a$, for all a, b in $\mathscr{R}$ (that is, $+$ is commutative).

(*iii*) There is an element $0_{\mathscr{R}}$ in $\mathscr{R}$ such that $a + 0 = a$, for all a in $\mathscr{R}$ (that is, $0_{\mathscr{R}}$ is the additive identity).

2. $\mathscr{R}$ is a monoid under multiplication, meaning that:

   (*i*) $(a \cdot b) \cdot c = a \cdot (b \cdot c)$, for all a, b, c in $\mathscr{R}$ (that is, $\cdot$ is associative).

   (*ii*) There is an element $1_{\mathscr{R}}$ in $\mathscr{R}$ such that $a \cdot 1_{\mathscr{R}} = a$ and $1_{\mathscr{R}} \cdot a = a$, for all a in $\mathscr{R}$ (that is, $1_{\mathscr{R}}$ is the multiplicative identity).

3. Multiplication is distributive over addition, meaning that:

   (*i*) $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ for all a, b, c in $\mathscr{R}$ (left distributivity).

   (*ii*) $(b + c) \cdot a = (b \cdot a) + (c \cdot a)$ for all a, b, c in $\mathscr{R}$ (right distributivity).

Example 1.6.3. *1. It is easy to verify that if* $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ *then, the inverse of* A*, denoted* A^{-1}*, is* $A^{-1} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}$*. Besides, if* $B = \begin{pmatrix} 1 & 2 \\ 3 & \frac{4}{5} \end{pmatrix}$ *then, the inverse of* B*, denoted* B^{-1}*, is* $B^{-1} = \begin{pmatrix} -\frac{2}{13} & \frac{5}{13} \\ \frac{15}{26} & -\frac{5}{26} \end{pmatrix}$.

2. Define A *the element of* $\mathscr{M}_2(\mathbf{K})$ *by setting:* $A := \begin{pmatrix} a & b \\ c & d \end{pmatrix}$*. We will show, during recitations that* A *is invertible if and only if* $ad - bc \neq 0$*. When this is the cas, we have the equality:*

$$A^{-1} = \frac{1}{ad - bc} \cdot \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Remark 1.6.4. *1. If* A *is invertible and if* $G \cdot A = I$*, then* $G = GAA^{-1} = A^{-1}$*. Likewise, if* $AD = I$*, then* $D = A^{-1}$*. In particular, the inverse of a matrix, when it exists, is unique.*

2. If there are two matrices G *and* D *such that* $GA = AD = I$*, then we have* $G = GAD = D$ *and* A *is invertible and its inverse is* $A^{-1} = G = D$*.*

3. We will see later that, if there exists a matrix G *such that* $GA = I$*, then* A *is an invertible matrix and* $A^{-1} = G$*.*

Proposition 1.6.5. *Let* A *and* B *be two matrices in* $\mathbf{GL}_p(\mathbf{K})$ *then:*

1. AB *is an invertible matrix and its inverse is* $B^{-1}A^{-1}$*. In other words, the inverse of a product of matrices is the product of the inverses but in reverse order.*

2. ${}^tA$ *is invertible and its inverse is* ${}^t(A^{-1})$*.*

Proof. Will be proved during recitations. □

Example 1.6.6. *Let* A *and* B *be the two matrices defined at Example 1.6.3. It is easy to check that:*

$$(A \cdot B)^{-1} = \begin{pmatrix} 7 & \frac{18}{5} \\ 15 & \frac{46}{5} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{23}{26} & -\frac{9}{26} \\ -\frac{75}{52} & \frac{35}{52} \end{pmatrix}$$

while

$$B^{-1} \cdot A^{-1} = \begin{pmatrix} -\frac{2}{13} & \frac{5}{13} \\ \frac{15}{26} & -\frac{5}{26} \end{pmatrix} \cdot \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{23}{26} & -\frac{9}{26} \\ -\frac{75}{52} & \frac{35}{52} \end{pmatrix}.$$

Chapter 2

Linear Systems

2.1 Linear System as a Matrix Equation

A system with p linear equations with q unknowns is of the form:

$$(\mathcal{S}) \qquad \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots a_{1q}x_q = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots a_{2q}x_q = b_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{p1}x_1 + a_{p2}x_2 + \cdots a_{pq}x_q = b_p \end{cases}$$

where a_{ij} and b_j are elements of $\mathbf{K}$ and the x_i are the unknowns that we want to determine. A solution of the system is a q-tuple denoted $(x_1, x_2, \cdots, x_q)$ of $\mathbf{K}^q$ for which the p equalities are satisfied. We are now interested to the the set of all solutions of the system. Let's introduce the following notations:

$$A := (a_{i,j}), \qquad X := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{pmatrix}, \qquad \& \qquad B := \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix}.$$

Using the notations defined, it is clear that $(\mathcal{S})$ can be written as

$$(\mathcal{S}') \qquad AX = B.$$

Matrix A is called *matrix of the system and the column matrix B is called second member of the system.* When $B = (0)$ (*i.e.* $b_i = 0$, for all $1 \leqslant i \leqslant p$), the system is said to be *homogeneous.* In this case it has an obvious solution that is the null solution $(0, \cdots, 0)$. This latter is also called trivial solution.

Definition 2.1.1 (Equivalent Systems). *Two linear systems of equations, which have the same number of unknowns are said to be equivalent if they both have the same set of solutions.*

Remark 2.1.2. *Two equivalent linear systems have the same number of unknowns but they might have different numbers of equations. Hence the two systems:*

$$(\mathcal{S}_1) : \begin{cases} x + y = 0 \\ 2x + 2y = 0 \end{cases} \qquad \text{and} \qquad (\mathcal{S}_2) : \left\{ \; x + y = 0 \right.$$

are equivalent.

Proposition 2.1.3. *Let A be a $p \times q$ matrix and let G be a $p \times p$ invertible matrix. The systems $AX = B$ and $GAX = GB$ are equivalent.*

Proposition 2.1.4. *Let A be a $p \times q$ matrix and let G be a $p \times p$ invertible matrix. The systems $AX = B$ and $GAX = GB$ are equivalent.*

Proof. To be filled! See [CPY96, p.19 Proposition I.10] □

Remark 2.1.5. *In the particular case where A is an invertible matrix (i.e. A belongs to in $\mathbf{GL}_p(\mathbf{K})$), then this system is equivalent to the system $X = A^{-1}B$, which has a unique solution $((A^{-1}B)_1, (A^{-1}B)_2, \cdots, (A^{-1}B)_p)$. Such a system is called a **Cramer** (or **regular**) system.*

2.2 Elementary Operations

Let's translate, in term of matrix operations, the method of the Gaussian elimination for solving linear systems.
In this method we only use three elementary operations, which transform the initial system in an equivalent system (which therefore has the same solutions as the initial one). These operations are:

1. **Multiplication of an equation by a scalar.**

More precisely, denote $\boldsymbol{\mathscr{M}(\lambda, i)}$ the multiplication of the i^{th} equation by the scalar λ (which is assumed to be different from 0). Note that this means multiply the i^{th} row of the matrix A by the scalar λ.

2. **Interchange two equations of a system.**

More precisely, denote $\boldsymbol{\mathscr{P}(i, j)}$ the operation which consists of interchanging the i^{th} and j^{th} equations. Note that this means interchanging rows i and j of the matrix with each other.

3. **Adding an equation to another.**

More precisely, denote $\boldsymbol{\mathscr{S}(i, j)}$ the operation which consists of adding the j^{th} equation to the i^{th} one. Note that this means exactly adding the j^{th} row of A to the i^{th} one.

As we will see each of these operations is nothing but multiplying the matrix A of the system $(\mathscr{S}')$ by an invertible matrix, which is said to be associated to the elementary operation. As we will notice further, the matrix associated to any of the elementary operations described above can be obtained by applying this elementary operation to the unit matrix.

1. **Multiplication of a row by a scalar.**

Let λ be a positive number. Consider the $p \times p$ matrix defined by:

$$M(\lambda, i) := \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & & & \vdots \\ \vdots & \ddots & 1 & \ddots & & & \vdots \\ \vdots & & \ddots & \lambda & \ddots & & \vdots \\ \vdots & & & \ddots & 1 & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix} \begin{matrix} \\ \\ \\ \leftarrow i^{\text{th}}\text{row} \\ \\ \\ \\ \end{matrix}$$

Let A be a matrix with p rows. We easily see that the matrix $M(\lambda, i) \cdot A$ has the same rows as A except for the i^{th} row, for which all coefficients has been multiplied by λ.
This means that, applying the operation $\boldsymbol{\mathscr{M}(\lambda, i)}$ to the system $AX = B$ gives us the system

$$M(\lambda, i) \cdot AX = M(\lambda, i) \cdot B.$$

It is also easy to see that the operation $\boldsymbol{\mathscr{M}(\frac{1}{\lambda}, i)}$ "cancels" the effect of $\boldsymbol{\mathscr{M}(\lambda, i)}$. The consequence of this fact is that

$$M(\lambda, i) \cdot M\left(\frac{1}{\lambda}, i\right) = I.$$

In other words, for all λ in $\mathbf{R}^*$, $M(\lambda, i)$ is invertible and:

$$M(\lambda, i)^{-1} = M\left(\frac{1}{\lambda}, i\right).$$

2. **Interchanging two rows of a Matrix with each other.**

Consider the $p \times p$ matrix defined by:

$$P(i,j) := \begin{pmatrix}
1 & 0 & \cdots & 0 & & \cdots & & 0 & & \cdots & 0 \\
0 & \ddots & & \vdots & & & & \vdots & & & \vdots \\
\vdots & & 1 & & & & & 0 & & \cdots & 0 \\
0 & \cdots & & 0 & & \cdots & 0 & 1 & 0 & \cdots & 0 \\
 & & & & 1 & & & 0 & & & \\
\vdots & & & \vdots & & \ddots & & \vdots & & & \vdots \\
 & & & 0 & & & 1 & & & & \\
0 & \cdots & 0 & 1 & 0 & \cdots & & 0 & & \cdots & 0 \\
 & & & 0 & & & & & 1 & & \vdots \\
\vdots & & & \vdots & & & & \vdots & & \ddots & 0 \\
0 & \cdots & & 0 & & \cdots & & 0 & \cdots & 0 & 1
\end{pmatrix}
\begin{array}{l}
\\ \\ \\ \leftarrow i^{th} \text{ row} \\ \\ \\ \\ \leftarrow j^{th} \text{ row} \\ \\ \\ \\
\end{array}$$

Let A be a matrix with p rows. We easily see that the matrix $P(i,j) \cdot A$ has the same rows as A except for that the rows i and j have been interchanged.
This means that, applying the operation $\boldsymbol{\mathscr{P}(i,j)}$ to the system $AX = B$ gives us the system

$$P(i,j) \cdot AX = P(i,j) \cdot B.$$

It is also easy to see that the operation $\boldsymbol{\mathscr{P}(j,i)}$ "cancels" the effect of $\boldsymbol{\mathscr{P}(i,j)}$. The consequence of this fact is that

$$P(i,j) \cdot P(j,i) = I.$$

In other words, for all (i,j) in $[\![1,p]\!]^2$, $P(i,j)$ is invertible and:

$$P(i,j)^{-1} = P(j,i).$$

The $P(i,j)$ matrices are called permutation matrices.

3. **Adding an row of a Matrix to another one.**

Consider the $p \times p$ matrix [1] defined by:

$$S(i,j) := \begin{pmatrix} 1 & 0 & \cdots & 0 & & \cdots & & 0 & & \cdots & 0 \\ 0 & \ddots & & \vdots & & & & \vdots & & & \vdots \\ \vdots & & & 0 & & & & & & & \\ 0 & \cdots & 0 & 1 & 0 & \cdots & & 0 & & \cdots & 0 \\ & & & 0 & & & & & & & \\ \vdots & & & \vdots & & \ddots & & \vdots & & & \vdots \\ & & & 0 & & & & 0 & & & \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ & & & 0 & & & & 0 & & & \vdots \\ \vdots & & & \vdots & & & & \vdots & & \ddots & 0 \\ 0 & \cdots & & 0 & & \cdots & & 0 & \cdots & 0 & 1 \end{pmatrix} \begin{matrix} \\ \\ \\ \leftarrow i^{th} \text{ row} \\ \\ \\ \\ \leftarrow j^{th} \text{ row} \\ \\ \\ \\ \end{matrix}$$

Let A be a matrix with p rows. We easily see that the matrix $S(i,j) \cdot A$ is obtained from A by replacing its j^{th} row by the sum of the i^{th} and j^{th} rows. applying operation $\boldsymbol{\mathcal{S}(i,j)}$ to the system $AX = B$ gives us the system

$$S(i,j) \cdot AX = S(i,j) \cdot B.$$

"Canceling" operation $\boldsymbol{\mathcal{S}(i,j)}$ can be done by subtracting the i^{th} row of the new matrix (*i.e.* $S(i,j) \cdot A$) to its j^{th} row. More precisely we can do that in three steps.

(i) Changing the sign of the i^{th} row.

(ii) Add the new i^{th} row to the j^{th} row

(iii) Changing again the sign of the i^{th} row.

In other words, for all (i,j) in $[\![1,p]\!]^2$, the matrix $S(i,j)$ is invertible and:

$$S(i,j)^{-1} = M(-1,i) \cdot S(j,i) \cdot M(-1,i).$$

4. Generalization of the three elementary operations.

One can generalize the three previous operations. Indeed, for any non zero real λ,

$$S(\lambda, i, j) := M\left(\frac{1}{\lambda}, i\right) \cdot S(j,i) \cdot M(\lambda, i).$$

corresponds to the operation:"adding to the j^{th} row the i^{th} row multiplied by λ". We denote $\boldsymbol{S(\lambda, i, j)}$ this operation.

Remark 2.2.1. *Note that the matrices of these successive operations are written from the right to the left since, every time, one multiplies from the left hand side.*

One easily sees that $S(\lambda, i, j)$ is an invertible matrix. Moreover, one has:

$$S(\lambda, i, j)^{-1} := M\left(-\frac{1}{\lambda}, i\right) \cdot S(j,i) \cdot M(-\lambda, i) = S(-\lambda, i, j).$$

[1]We here chose $i < j$. The reader will write the case where $i > j$.

2.3 Gaussian Elimination & Row Reduced Echelon Matrices

Definition 2.3.1 (Row Echelon Matrix). *A $p \times q$ matrix R is said to be in row echelon form if it has the following "staircase" structure:*

$$R = \begin{pmatrix}
\circledast & * & \cdots & * & * & \cdots & * & * & \cdots & \cdots & * & * & * & \cdots & * \\
0 & 0 & \cdots & 0 & \circledast & \cdots & * & * & \cdots & \cdots & * & * & * & \cdots & * \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \circledast & \cdots & \cdots & * & * & * & \cdots & * \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & & \vdots & \vdots & & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 & \circledast & * & \cdots & * \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \cdots & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & 0
\end{pmatrix}.$$

The entries indicated by $\circledast$ are the pivots, and must be nonzero. The first r rows of R each contain exactly one pivot, but not all columns are required to include a pivot entry. The entries below the "staircase", indicated by the solid line, are all zero, while the non-pivot entries above the staircase, indicated by stars, can be anything. The last $p - r$ rows are identically zero and do not contain any pivots. there may, in exceptional situations, be one or more all zero initial columns.

Example 2.3.2. *The matrix*

$$R := \begin{pmatrix}
3 & 1 & 0 & 4 & 5 & -7 \\
0 & -1 & -2 & 1 & 8 & 0 \\
0 & 0 & 0 & 0 & 2 & -4 \\
0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}$$

is in row echelon form. It has three pivots that are: 3, -1 and 2.
Note moreover that the row echelon matrix can have several initial columns consisting of all zeros. As the exemple below shows.

$$S := \begin{pmatrix}
0 & 0 & 3 & 5 & -2 & 0 \\
0 & 0 & 0 & 0 & 5 & 3 \\
0 & 0 & 0 & 0 & 0 & -7
\end{pmatrix}.$$

This matrix is also in row echelon form. It has three pivots that are: 3, 5 and -7.

Definition 2.3.3 (Reduced Row Echelon Matrix). *A $p \times q$ matrix R is said to be in reduced row echelon form if it has the "staircase" structure given at Definition 2.3.1 and if all the pivots equal 1 and if all the elements above pivots, (in the columns which contains a pivot), equal 0. In this latter case, we call pivot columns the columns which contain the pivots.*

Example 2.3.4. *The matrix*

$$T := \begin{pmatrix} 0 & 0 & \mathbf{1} & * & * & * & \mathbf{0} & * & \mathbf{0} & * & * \\ 0 & 0 & \mathbf{0} & 0 & 0 & 0 & \mathbf{1} & * & \mathbf{0} & * & * \\ 0 & 0 & \mathbf{0} & 0 & 0 & 0 & \mathbf{0} & 0 & \mathbf{1} & * & * \\ 0 & 0 & \mathbf{0} & 0 & 0 & 0 & \mathbf{0} & 0 & \mathbf{0} & 0 & 0 \\ 0 & 0 & \mathbf{0} & 0 & 0 & 0 & \mathbf{0} & 0 & \mathbf{0} & 0 & 0 \end{pmatrix}$$

is in reduced row echelon form. Moreover, the pivot columns are here highlighted in bold font.

Definition 2.3.5 (Rank of a Matrix). *A $p \times q$ matrix R is said to be of rank r (where $0 \leqslant r \leqslant \min\{p, q\}$) if it can be obtained, from the $r \times r$ unit matrix, from one hand by adding $p - r$ rows of zeros at the bottom of the matrix and, from the other hand, by adding between columns j and $j+1$ (for j from 0 to r) as many number of columns of the form*

$$\begin{array}{cl} \begin{pmatrix} * \\ \vdots \\ * \\ 0 \\ \vdots \\ 0 \end{pmatrix} & \begin{array}{l} \\ \\ \leftarrow j^{th} \text{row} \\ \\ \\ \\ \end{array} \end{array}$$

where $*$ denotes any element of $\mathbf{K}$.

Example 2.3.6. *The rank of the matrix T, given at Example 2.3.4, is 3.*

We hence proved the following result, the roof of which is obvious.

Proposition 2.3.7. *The rank of a reduced row echelon matrix is its number of pivots.*

Remark 2.3.8. *In the particular case where R is reduced row echelon matrix with rank 0, we see that one can only add columns full of zeros. Thus all the coefficients of R are zeros. Thus $R = (0)$.*

If it is easy to determine the rank of a reduced row echelon matrix. However, one needs a concrete way to determine the rank of any matrix. This can be done thanks to the following result, the proof of which gives a concrete way to do so.

Theorem 2.3.9 (Gauss)**.** *Let A be a matrix in $\mathscr{M}_{p,q}(\mathbf{K})$. There exists an invertible matrix in $\mathscr{M}_p(\mathbf{K})$, denoted G, such that GA is a reduced row echelon matrix.*

Note that the proof of this result, given below, is known as **Gaussian Elimination**.

Proof. The Matrix G will be built as a product of matrices associated to the elementary operations described in Section 2.2. In view of Proposition 1.6.5, Matrix G will be invertible since it is a product of invertible matrices.
We therefore just have to show that, by a succession of elementary operations, one can transform Matrix $A := (a_{i,j})_{(i,j) \in [\![1,p]\!] \times [\![1,q]\!]}$ into a reduced row echelon matrix.

1. **Reduction of the first column of A**

Assume that the the first column of A has at least one non zero coefficient, say $a_{k,1}$. By multiplying row k by $\frac{1}{a_{k1}}$ we now are in the case where the first column of A has one coefficient, namely a_{k1}, equals to 1 exactly. By interchanging rows 1 and k, one can reduce our reasoning to the case where this coefficient (equal to 1) is on the first row of A.
If we add successively, for i from 2 to p, the first row multiplied by $-a_{i1}$ to the i^{th} row, we get a matrix the first column of which is full of zeros except for the first element which equals 1. One can summarize this process by writing:

$$\begin{pmatrix} a_{11} & \cdots \\ a_{21} & \\ \vdots & \\ a_{i1} & \cdots \\ \vdots & \\ a_{p1} & \cdots \end{pmatrix} \xrightarrow{\mathscr{M}\left(\frac{1}{a_{i1}}, i\right)} \begin{pmatrix} a_{11} & \cdots \\ a_{21} & \\ \vdots & \\ 1 & \cdots \\ \vdots & \\ a_{p1} & \cdots \end{pmatrix} \xrightarrow{\mathscr{P}(1,i)} \begin{pmatrix} 1 & \cdots \\ a'_{21} & \\ \vdots & \\ \vdots & \cdots \\ \vdots & \\ a'_{p1} & \cdots \end{pmatrix}$$

$$\xrightarrow{\mathscr{S}\left(-a'_{21}, 1, 2\right)} \begin{pmatrix} 1 & \cdots \\ 0 & \\ a'_{31} & \\ \vdots & \cdots \\ \vdots & \\ a'_{p1} & \cdots \end{pmatrix} \longrightarrow \cdots \xrightarrow{\mathscr{S}\left(-a'_{p1}, 1, p\right)} \begin{pmatrix} 1 & \cdots \\ 0 & \\ 0 & \\ 0 & \cdots \\ 0 & \\ 0 & \cdots \end{pmatrix}$$

In other words, in order to go from a matrix A, the first column of which was $\begin{pmatrix} a_{11} & \cdots \\ a_{21} & \\ \vdots & \\ a_{i1} & \cdots \\ \vdots & \\ a_{p1} & \cdots \end{pmatrix}$ to a matrix the first column of which is $\begin{pmatrix} 1 & \cdots \\ 0 & \cdots \\ \vdots & \\ 0 & \\ \vdots & \\ 0 & \cdots \end{pmatrix}$ we multiplied A, from the left hand side, by the matrix

$$G := S(-a'_{p1}, 1, p) \cdots S(-a'_{21}, 1, 2) \cdot P(1, i) \cdot M\left(\frac{1}{a_{i1}}, i\right).$$

2. **Reduction of the other columns of A**

In order to prove the theorem we do an induction on p. The first part of the demonstration proved that, by elementary operations on the rows of A, can get a matrix the first column of which is of the form:

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

- If $q = 1$ (or if $p = 1$)

then this ends the proof and the matrix we obtain is in reduced row echelon form. Moreover it has a rank equal to 0 or 1).

- Assume that If $q \geqslant 2$ and that the theorem is proved up to the order $q-1$.

Once we reduced the first column (as we did in **1**), one has to get the reduced row echelon form of a matrix, denoted A_1, which is of one of these types

$$\begin{pmatrix} 0 & \\ \vdots & \boldsymbol{B} \\ 0 & \end{pmatrix} \qquad \begin{pmatrix} 1 & * & \cdots & * \\ 0 & & & \\ \vdots & & \boldsymbol{B} & \\ 0 & & & \end{pmatrix}$$

where $\boldsymbol{B}$ is a matrix with $q-1$ columns. According to the Induction Assumption, one can transform, using only elementary operations on the rows, the matrix $\boldsymbol{B}$ in a reduced row echelon matrix.
By applying these elementary operations on the matrix A_1 (taking into account the shifting between line numbering of $\boldsymbol{B}$ and A_1 (in the second case), we get a matrix A_2 which has one of the form below, but in which $\boldsymbol{B}$ is reduced.
In the first case, we are done since by adding a column full of zeros at the left of a reduced row echelon matrix we still get a reduced row echelon matrix. In the second case, we just have to make appear zeros on the first row of the matrix A_2 and only for the pivot columns of matrix $\boldsymbol{B}$. Denote j the index of one of these columns. This column is of the form

$$\begin{array}{cl} \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} & \begin{array}{l} \\ \\ \\ \\ \leftarrow j^{\text{th}}\text{row} \\ \\ \\ \\ \end{array} \end{array}$$

We therefore only have to add to the first row of A_2, the j^{th} row of A_2 multiplied by $-\alpha$ in order to get the desired form for the pivot column. This does not affect the columns of A_2 which extend the pivot columns of $\boldsymbol{B}$ which have a different index from j (but modify in general the other columns). This achieves the proof of the theorem. $\square$

Remark 2.3.10. *Let A be in $\mathscr{M}_{p,q}(\mathbf{K})$. In order to determine the matrices G and $R := GA$, one can use Gaussian Elimination. In order to avoid having to remember the elementary operations we have done to reach this goal, we use the fact that G is obtained by performing all these elementary operations on the unit matrix. Thus we are going to perform the required elementary operations on the matrix A and, in parallel, on the unit matrix $I_{\min\{p,q\}}$. Once the matrix A has been transformed into a reduced row echelon matrix R, $I_{\min\{p,q\}}$ will have been transformed into the matrix G.*

Example 2.3.11. *Give the row reduced echelon form of the matrix $A := \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$. We start with the following position:*

Starting positon	$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
Operation $\mathscr{S}(-4, 1, 2)$	$\begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ -4 & 1 \end{pmatrix}$
Operation $\mathscr{M}(-\frac{1}{3}, 2)$	$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 4/3 & -1/3 \end{pmatrix}$
Operation $\mathscr{S}(-2, 2, 1)$	$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{pmatrix}$	$\begin{pmatrix} -5/3 & 2/3 \\ 4/3 & -1/3 \end{pmatrix}$

The matrix $R := \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{pmatrix}$ is in reduced row echelon form and $G := \begin{pmatrix} -5/3 & 2/3 \\ 4/3 & -1/3 \end{pmatrix}$ is the invertible matrix such that $GA = R$.

Remark 2.3.12. *One can use a lighter and more instinctive way to denote the elementary operations we are doing when we try to determine the reduced row echelon form. Namely $R_i \longleftarrow R_i - 2R_j$ means that the new row R_i is the former one from which we subtracted 2 times R_j. Hence the previous example can be rewritten as follow.*

Starting positon	$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
$R_2 \longleftarrow R_2 - 4R_1$	$\begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ -4 & 1 \end{pmatrix}$
$R_2 \longleftarrow -R_2/3$	$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 4/3 & -1/3 \end{pmatrix}$
$R_1 \longleftarrow R_1 - 2R_2$	$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{pmatrix}$	$\begin{pmatrix} -5/3 & 2/3 \\ 4/3 & -1/3 \end{pmatrix}$

2.4 Application of Gaussian Elimination to inversion of Matrices

We will focus, in this section, on **square** matrices.

We saw in the previous section that if the reduced row echelon matrix associated to a **square** matrix A by the Gauss Elimination Theorem is the identity matrix then A is an invertible matrix. The Gauss Elimination then allows not only to establish that a matrix A is invertible, by showing that $R = I$, but also to compute its inverse G (since $GA = R = I$).

Example 2.4.1. *Let us show that* $A := \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix}$ *is invertible and let us compute its inverse.*

$$\text{Starting positon} \quad \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{array}{l} R_2 \longleftarrow R_2 - 4R_1 \\ R_3 \longleftarrow R_3 - 7R_1 \end{array} \quad \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -11 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ -7 & 0 & 1 \end{pmatrix}$$

$$R_3 \longleftarrow R_3 - 2R_2 \quad \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ 1 & -2 & 1 \end{pmatrix}$$

$$R_2 \longleftarrow -R_2/3 \quad \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 4/3 & -1/3 & 0 \\ 1 & -2 & 1 \end{pmatrix}$$

$$R_1 \longleftarrow R_1 - 2R_2 \quad \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} -5/3 & 2/3 & 0 \\ 4/3 & -1/3 & 0 \\ 1 & -2 & 1 \end{pmatrix}$$

$$\begin{array}{l} R_1 \longleftarrow R_1 + R_3 \\ R_2 \longleftarrow R_2 - 2R_3 \end{array} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} -2/3 & -4/3 & 1 \\ -2/3 & 11/3 & -2 \\ 1 & -2 & 1 \end{pmatrix}$$

Since the reduced row echelon matrix is the unit matrix, it appears that A is invertible and that its inverse if

$$A^{-1} = \begin{pmatrix} -2/3 & -4/3 & 1 \\ -2/3 & 11/3 & -2 \\ 1 & -2 & 1 \end{pmatrix}.$$

We can therefore check that $\begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -11 \end{pmatrix} \cdot \begin{pmatrix} -2/3 & -4/3 & 1 \\ -2/3 & 11/3 & -2 \\ 1 & -2 & 1 \end{pmatrix} = I_3.$

2.5 Application of Gaussian Elimination to LU Factorization

We will focus, in this section again, on **square** matrices.

Introduced by Polish mathematician Tadeusz Banachiewicz in 1938, lower–upper LU decomposition or factorization factors a matrix as the product of a lower triangular matrix and an upper triangular matrix. Solving a triangular [2] linear system is easy and easier than solving a non triangular linear system. We therefore understand that it might be of interest to transform the research of the solution of a given linear system by the research of the solution of triangular systems. This is one of the interests of the LU Factorization. As we will se at the end of this section, this technique is very useful in numerical analysis and in linear algebra. Indeed, computers usually solve square systems of linear equations using LU decomposition, and it is also a key step when inverting a matrix or computing the determinant of a matrix.

Definition 2.5.1 (**LU Factorization).** *Let A be in $\mathscr{M}_p(\mathbf{K})$. A presentation $A = LU$, in which L belongs to $\mathscr{M}_p(\mathbf{K})$ and is a lower triangular and U belongs to $\mathscr{M}_p(\mathbf{K})$ and is upper triangular, is called an LU factorization of A.*

As we will see later, not all square matrices admit an LU factorization. Hence $B := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ does not admit such a decomposition. It is therefore important to be able to know which sets of square matrices do admit such a decomposition. Let us start by a definition.

Definition 2.5.2 (Regular Gaussian Elimination for Square Matrices). *We say that a Gaussian elimination is **regular** if the Gaussian elimination algorithm successfully reduces the given matrix A to a row echelon matrix:*

1. *only by the elementary operations described at Section 2.2, **at the exception of interchanging rows**.*
2. *the pivots obtained via the Gaussian elimination algorithm are all non-zero pivots.*

By extension one can define **regular** square matrices.

Definition 2.5.3 (Regular Square Matrices). *We say that a square matrix is **regular** if the Gaussian Elimination Algorithm to get the row echelon form of A successfully reduces to upper triangular form U with all non-zero pivots and without any interchanging of rows.*

Example 2.5.4. *Define the following matrices*

$$A := \begin{pmatrix} 2 & -3 & 1 \\ -2 & 2 & -3 \\ 4 & -9 & 2 \end{pmatrix}, \qquad B := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad C := \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix},$$

It is easy to check that A is regular, while neither B nor C is.

LU Factorization of Regular Matrices

Regular matrices have a LU factorization. Moreover one can be accurate on the value of the pivots of both L and U. This is the object of the next result.

[2] *i.e.* a linear system for which Matrix A is triangular

Theorem 2.5.5 (Regular LU Factorization). *A matrix A is **regular** if and only if it can be factored*

$$A = LU \tag{1}$$

where L is a lower unitriangular matrix, having all $\mathbf{1}$'s on the diagonal, and U is upper triangular with non zero diagonal entries, which are the pivots of A. The non zero off-diagonal entries l_{ij} for $i > j$ appearing in L prescribe the elementary row operations that bring A into upper triangular form; namely, one subtracts $l_{i,j}$ times row j from row i at the appropriate step of the Gaussian Elimination process.

Remark 2.5.6. *In order to avoid any confusion between LU factorization (given at Definition 2.5.1, and where we do not have any information about the diagonal entries of U) and the LU factorization described right above (where we know that matrix U has non zero diagonal entries), we will call this latter **regular** LU **factorization**.*

In practice, to find the **regular** LU factorization of a square matrix A, one applies the regular Gaussian Elimination algorithm to reduce A to its upper triangular form U. The entries of L can be filled in during the the course of the calculation with the negatives of the multiples used in the elementary row operations. If the algorithm fails to be completed, which happens whenever zero appears in any diagonal pivot position or if we have to interchange rows to get the row echelon form of A, then the original matrix does not have an **regular** LU factorization.

Example 2.5.7. *[OS18, Ex. 1.4, p.18] Let's compute the LU factorization of the matrix*

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

At this stage we do not even know if such a factorization does exist for A. Only the Gaussian Elimination algorithm will tell us if it does. Thus, applying the Gaussian Elimination algorithm, we begin by defining A_1 to be A and L_1 to be I_3. We hence start with:

$$\begin{pmatrix} 2 & 1 & 1 \\ 4 & 5 & 2 \\ 2 & -2 & 0 \end{pmatrix} := A_1 \qquad \begin{pmatrix} \mathbf{1} & 0 & 0 \\ 0 & \mathbf{1} & 0 \\ 0 & 0 & \mathbf{1} \end{pmatrix} := L_1$$

$$\begin{array}{l} R_2 \longleftarrow R_2 - \mathbf{2}R_1 \\ R_3 \longleftarrow R_3 - \mathbf{1}R_1 \end{array} \begin{pmatrix} 2 & 1 & 1 \\ 0 & 3 & 0 \\ 0 & -3 & -1 \end{pmatrix} := A_2 \qquad \begin{pmatrix} \mathbf{1} & 0 & 0 \\ \mathbf{2} & \mathbf{1} & 0 \\ \mathbf{1} & 0 & \mathbf{1} \end{pmatrix} := L_2$$

$$R_3 \longleftarrow R_3 + \mathbf{1}R_2 \quad \begin{pmatrix} 2 & 1 & 1 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \end{pmatrix} := A_3 \qquad \begin{pmatrix} \mathbf{1} & 0 & 0 \\ \mathbf{2} & \mathbf{1} & 0 \\ \mathbf{1} & \mathbf{-1} & \mathbf{1} \end{pmatrix} := L_3$$

Note that A_3 is an upper triangular matrix while L_3 is a lower triangular matrix. Thus one defines U to be A_3 and L to be L_3. In other words, define matrices L and U by setting:

$$L := \begin{pmatrix} \mathbf{1} & 0 & 0 \\ \mathbf{2} & \mathbf{1} & 0 \\ \mathbf{1} & \mathbf{-1} & \mathbf{1} \end{pmatrix} \qquad \& \qquad U := \begin{pmatrix} 2 & 1 & 1 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

*Note that the diagonal entries of U are the pivots of A. Since they are all non zero here, it is clear, in view of Theorem 2.5.5, that A is regular. Therefore the LU factorization we just gave is the **regular** LU factorization of A.*
Note that L has always all its diagonal elements equal to 1 while its entries lying below the main diagonal are the negatives of the multiples we used during the elimination procedure. Moreover, it is easy to check that

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

Remark 2.5.8. *By trying Gaussian elimination on matrices B and C given in Example 2.5.4 we immediately see that the algorithm requires an interchange of rows. Thus none of these two matrices is regular. As a consequence none of them admit a **regular** LU factorization.*

As the next subsection will show, it is possible, if we get the rid of the condition concerning the interchanging of rows, to extend the **regular** LU factorization to any invertible matrix (and in fact to a bigger set of matrices). The price to pay will be that we will not have $A = LU$ anymore but $PA = LU$, for a certain invertible matrix P.

$PA = LU$ Factorization

Let us start with the most general result, the proof of which can be found in [HJ13, Theorem 3.5.8].

Theorem 2.5.9 ($PA = LU$ **Factorization).** *For each matrix A in $\mathscr{M}_p(\mathbf{K})$ there is a permutation matrix P in $\mathscr{M}_p(\mathbf{K})$, a unit lower triangular L in $\mathscr{M}_p(\mathbf{K})$, and an upper triangular U in $\mathscr{M}_p(\mathbf{K})$ such that $PA = LU$. We call **permuted** LU **factorization** such a decomposition.*

Remark 2.5.10. *A unit lower triangular matrix is just a lower triangular matrix the diagonal entries of which are all equal to* $\mathbf{1}$.

In other words, the previous result states that permuted LU factorization is always possible for a square matrix. However, if one wants that the pivots of U are all non zero one must restrain the set of square matrices considered. More precisely, let us first define Nonsingular Matrix.

Definition 2.5.11 (Nonsingular Matrix). *We say that a square matrix A is **nonsingular** if the Gaussian Elimination Algorithm successfully reduces it to an upper triangular matrix U with all non-zero pivots (on the diagonal then) only by the elementary rows operations described at Section 2.2.*

Remark 2.5.12. *A matrix which is not nonsingular is said to be singular.*

Definition 2.5.13 (Nonsingular $PA = LU$ Factorization). *For any non singular matrix A, we call **nonsingular** $PA = LU$ Factorization or **nonsingular** permuted LU Factorization the $PA = LU$ Factorization granted by Theorem 2.5.9.*

Remark 2.5.14. *The main information added when one goes from $PA = LU$ Factorization to **nonsingular** $PA = LU$ Factorization is that, in this latter case all the diagonal entries of U are non zero.*

We have the following result.

Theorem 2.5.15 (Nonsingular $PA = LU$ Factorization). *Let A be a matrix in $\mathscr{M}_p(\mathbf{K})$. The following conditions are equivalent:*

1. *A is nonsingular.*
2. *A is invertible.*
3. *A has p nonzero pivots.*
4. *A admits a **nonsingular** permuted LU factorization: $PA = LU$.*

A practical method to construct a permuted LU factorization of a given matrix A, whether it is nonsingular or not, of $\mathscr{M}_p(\mathbf{K})$ would proceed as follows. First set up $P = L = I_p$. The matrix P will keep track of the permutations performed during the Gaussian Elimination process, while the entries of L below the diagonal are gradually replaced by the negatives of the multiples used in the corresponding elementary row operations (except the interchanging of two rows). Each time two rows of A are interchanged, the same two rows of P will be interchanged as well. Moreover, any pair of entries that both lie *below* in these same two rows of L must also be interchanged, while entries lying on and above its diagonal need to stay in their place. At a successful conclusion to the procedure, A will have been converted into the upper triangular matrix U, while L and P will assume their final form. Here is an illustrative example.

Example 2.5.16. *Let us give a permuted LU factorization of the matrix*

$$A := \begin{pmatrix} 1 & 2 & -1 & 0 \\ 2 & 4 & -2 & -1 \\ -3 & -5 & 6 & 1 \\ -1 & 2 & 8 & -2 \end{pmatrix}.$$

We hence start with:

$$\begin{pmatrix} 1 & 2 & -1 & 0 \\ 2 & 4 & -2 & -1 \\ -3 & -5 & 6 & 1 \\ -1 & 2 & 8 & -2 \end{pmatrix} := A_1 \qquad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} := L_1 \qquad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} := P_1$$

To begin the procedure, we eliminate the entries below the first pivot.

$$\begin{array}{r} R_2 \longleftarrow R_2 - \mathbf{2}R_1 \\ R_3 \longleftarrow R_3 + \mathbf{3}R_1 \\ R_4 \longleftarrow R_4 + \mathbf{1}R_1 \end{array} \begin{pmatrix} 1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 1 & 3 & 1 \\ 0 & 4 & 7 & -2 \end{pmatrix} := A_2 \quad \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 \\ \mathbf{2} & \mathbf{1} & 0 & 0 \\ \mathbf{-3} & 0 & \mathbf{1} & 0 \\ \mathbf{-1} & 0 & 0 & \mathbf{1} \end{pmatrix} := L_2 \quad \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{1} \end{pmatrix} := P_2.$$

Since the $(2, 2)$ entry of A_2 is zero, we interchange rows 2 and 3 which leads us to:

$$R_2 \longleftrightarrow R_3 \quad \begin{pmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 4 & 7 & -2 \end{pmatrix} := A_3 \quad \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 \\ \mathbf{-3} & \mathbf{1} & 0 & 0 \\ \mathbf{2} & 0 & \mathbf{1} & 0 \\ \mathbf{-1} & 0 & 0 & \mathbf{1} \end{pmatrix} := L_3 \quad \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 \\ 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} \end{pmatrix} := P_3.$$

We interchanged the same two rows of P, while in L we only interchanged the already computed entries in tis second and third rows that lie in its first column below the diagonal. We then eliminate the nonzero entry lying below the $(2,2)$ pivot, leading to:

$$\underset{R_4 \longleftarrow R_4 - 4R_2}{} \begin{pmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -5 & -6 \end{pmatrix} := A_4 \quad \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 \\ \mathbf{-3} & \mathbf{1} & 0 & 0 \\ \mathbf{2} & 0 & \mathbf{1} & 0 \\ \mathbf{-1} & \mathbf{4} & 0 & \mathbf{1} \end{pmatrix} := L_4 \quad \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 \\ 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} \end{pmatrix} := P_4.$$

A final row interchange places the matrix in upper triangular form:

$$\underset{R_3 \longleftrightarrow R_4}{} \begin{pmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & -5 & -6 \\ 0 & 0 & 0 & -1 \end{pmatrix} := A_5 \quad \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 \\ \mathbf{-3} & \mathbf{1} & 0 & 0 \\ \mathbf{-1} & \mathbf{4} & \mathbf{1} & 0 \\ \mathbf{2} & 0 & 0 & \mathbf{1} \end{pmatrix} := L_5 \quad \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{1} \\ 0 & \mathbf{1} & 0 & 0 \end{pmatrix} := P_5.$$

Again, we performed the same row interchange on P, while interchanging only the thirs and fourth row entries of L that lie below the diagonal. We thus define

$$U := A_5, \qquad L := L_5, \qquad \& \qquad P := P_5.$$

In other words, we have:

$$U := \begin{pmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & -5 & -6 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad L := \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 \\ \mathbf{-3} & \mathbf{1} & 0 & 0 \\ \mathbf{-1} & \mathbf{4} & \mathbf{1} & 0 \\ \mathbf{2} & 0 & 0 & \mathbf{1} \end{pmatrix} \; \& \; P := \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{1} \\ 0 & \mathbf{1} & 0 & 0 \end{pmatrix}.$$

U having all non zero pivots, it is, as well as A, non singular. Besides, one can verify that:

$$PA = \begin{pmatrix} 1 & 2 & -1 & 0 \\ -3 & -5 & 6 & 1 \\ -1 & 2 & 8 & -2 \\ 2 & 4 & -2 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -3 & -1 & 0 & 0 \\ -1 & 4 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & -5 & -6 \\ 0 & 0 & 0 & -1 \end{pmatrix} = LU. \tag{2}$$

Thus by rearranging the equations in the order first, third, fourth, second, as prescribed by P, we obtain an equivalent linear system whose coefficient matrix PA is regular, in accordance with Theorem 2.5.9.

Remark 2.5.17. ***1.*** *In order not to overload notations one usually does not specify the nature of the LU factorization (i.e. regular or nonsingular nor if it is or not a permuted one) since it is clear from the context.*

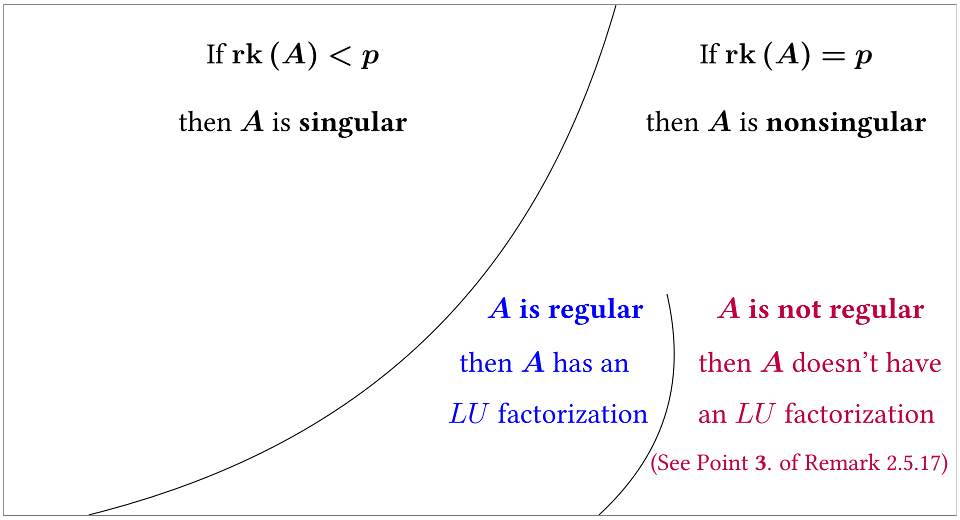
2. *Regular (if it exists) or permuted LU factorization is not, in general, unique.*

3. *Note that when one deals with a non invertible matrix A, the permuted LU factorization exists but there will be some zeros on the diagonal of U (as show Example 2.6.12 below).*

4. *The advantages of such a factorization will be given in Remark 2.6.10.*

Readers interested in more details about this factorization should refer to [HJ13]. The figure below gives a summary of how the regular, nonsingular and singular matrices are linked to each others.

Let A be in $\mathscr{M}_p(\mathbf{R})$, where $p \in \mathbf{N}^*$.



The set of square matrices $\mathscr{M}_p(\mathbf{K})$

2.6 Application of Gaussian Elimination to solving Linear System

Let us first state the general results that apply to all rectangular linear systems.

Results for general linear systems

Proposition 2.6.1. *A system of p linear equations with q unknowns has either 0, or 1 or infinitely many solutions. If $q > p$, it has either 0, or infinitely many solutions.*

Definition 2.6.2 (Rank of a general Matrix). *We define the rank of a $p \times q$ matrix as being the rank of the reduced row echelon matrix associated to A in the Gauss theorem. We denote it r.*
In particular, $0 \leqslant r \leqslant \min\{p, q\}$.

Proposition 2.6.3. *Let A be a $p \times q$ matrix and G be a $p \times p$ invertible matrix. Matrices A and B have the same rank.*

Definition 2.6.4 (Rank of a Linear System). *We define the rank of a linear system $AX = B$, as being the rank of the matrix A.*

Corollary 2.6.5. *An homogeneous linear system (i.e. with a second member equal to (0)) with p equations and q unknowns has a non zero solution if and only if the rank of its matrix is smaller than q. This is in particular the case if $p < q$.*

Results About Square Linear Systems

The results stated in the previous section remain valid. However, one can be more accurate when it comes to square linear systems.
From now on, define Let A be a matrix of $\mathscr{M}_p(\mathbf{K})$ and B be a column matrix with p rows. Denote $X := \begin{pmatrix} x_1 \\ x_1 \\ \vdots \\ x_p \end{pmatrix}$ and define

$$AX = B. \tag{$\mathscr{S}$}$$

The next result gives the methodology to solve such a linear system if A is invertible, using the inverse of A.

Theorem 2.6.6 (Characterisation of Invertible Matrices)**.** *Considering the System* ($\mathscr{S}$) *described above, the following conditions are equivalent:*

1. A *is an invertible matrix.*
2. *The rank of* A *equals* p *(in other words the reduced matrix, associated to* A *has no null row).*
3. *The system* $AX = (0)$ *has only one solution which is* $(0, \cdots, 0)$.
4. *For all column matrix* B*, the system* $AX = B$ *has at least one solution.*
5. *For all column matrix* B*, the system* $AX = B$ *has a unique solution.*

Moreover, when one of these conditions holds, then the reduced matrix, associated to A*, is* I *and the unique solution of the system* $AX = B$ *is* $((A^{-1}B)_1, (A^{-1}B)_2, \cdots, (A^{-1}B)_p)$.

Proof. To be filled! See [CPY96] □

Corollary 2.6.7. *Let* A *be a matrix in* $\mathscr{M}_p(\mathbf{K})$*. The following conditions are equivalent:*

1. A *is an invertible matrix.*
2. *There exists* G *in* $\mathscr{M}_p(\mathbf{K})$ *such that* $GA = I$.
3. *There exists* D *in* $\mathscr{M}_p(\mathbf{K})$ *such that* $AD = I$.

Proof. To be filled! See [CPY96, p.27] □

The next result gives the methodology to solve such a linear system if A is invertible, using the $PA = LU$ factorization.

Theorem 2.6.8 (Solving linear system with $PA = LU$ factorization)**.** *Let us consider the System* ($\mathscr{S}$) *described above. Assuming that* A *is an invertible matrix,* ($\mathscr{S}$) *admits a unique solution, that can be found using* $PA = LU$ *factorization.*

Methodology for solving square linear system using $PA = LU$ factorization

Once the permuted $PA = LU$ factorization is established, the solution to the original system $(\mathscr{S})$ is obtained by applying the Forward and back Substitution algorithm presented above. Explicitly, we first multiply the system $AX = B$ by P, from the left hand side, leading to:

$$PAX = PB =: B'. \tag{3}$$

Note that B' has been obtained by permuting the entries of B in the same fashion as the rows of A. Moreover (3) can be rewritten under the form:

$$LUX = B'. \tag{4}$$

We then solve the two triangular systems. The first one is

$$LC = B', \tag{5}$$

where C is the unknown vector in (5), and is solved by Forward Substitution. The second triangular system to solve, by Backward Substitution this time, is

$$UX = C \tag{6}$$

in the unknown X.

Example 2.6.9. *Let us solve the following linear system*

$$(\mathscr{T}) \qquad AX = B,$$

where

$$A := \begin{pmatrix} 1 & 2 & -1 & 0 \\ 2 & 4 & -2 & -1 \\ -3 & -5 & 6 & 1 \\ -1 & 2 & 8 & -2 \end{pmatrix} \qquad \textit{and} \qquad B := \begin{pmatrix} 1 \\ -1 \\ 3 \\ 0 \end{pmatrix}.$$

1. ***Using the inverse of*** A.

The Gauss elimination algorithm easily shows that A is invertible and that

$$A^{-1} = \begin{pmatrix} -\frac{12}{5} & \frac{32}{5} & \frac{18}{5} & -7/5 \\ 8/5 & -\frac{13}{5} & -7/5 & 3/5 \\ -1/5 & 6/5 & 4/5 & -1/5 \\ 2 & -1 & 0 & 0 \end{pmatrix}.$$

According to the conclusion of Theorem 2.6.6, we therefore know that $X = A^{-1}B$. Thus

$$X = \begin{pmatrix} (A^{-1}B)_1 \\ (A^{-1}B)_2 \\ (A^{-1}B)_3 \\ (A^{-1}B)_4 \end{pmatrix} = \begin{pmatrix} -\frac{12}{5} & \frac{32}{5} & \frac{18}{5} & -7/5 \\ 8/5 & -\frac{13}{5} & -7/5 & 3/5 \\ -1/5 & 6/5 & 4/5 & -1/5 \\ 2 & -1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 1 \\ 3 \end{pmatrix}. \tag{7}$$

Define $Sol_{\mathscr{T}}$ *the set of solutions of System* $\mathscr{T}$. *We just proved that:*

$$\boxed{Sol_{(\mathscr{T})} = \left\{{}^t(2,0,1,3)\right\},}$$

where ^{t} *is the transpose symbol.*

2. ***Using the*** $PA = LU$ ***factorization.***

The $PA = LU$ *factorization of* A *has been computed at* (2) *(on page 22) and is therefore not recalled here. According to the methodology given right after Theorem 2.6.8, define*

$$B' := \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{1} \\ 0 & \mathbf{1} & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 0 \\ -1 \end{pmatrix}.$$

Denote $C := \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix}$. *We know that we first need to solve, by forward substitution, the system*

$$LC = B', \tag{8}$$

in C. *Since* (8) *reads*

$$\begin{pmatrix} \mathbf{1} & 0 & 0 & 0 \\ \mathbf{-3} & \mathbf{1} & 0 & 0 \\ \mathbf{-1} & \mathbf{4} & \mathbf{1} & 0 \\ \mathbf{2} & 0 & 0 & \mathbf{1} \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 0 \\ -1 \end{pmatrix}, \tag{9}$$

we easily get the unique solution of (8), *which is:* $C = \begin{pmatrix} 1 \\ 6 \\ -23 \\ -3 \end{pmatrix}$. *One then has to solve, by backward substitution this time, the following system*

$$UX = C, \tag{10}$$

in the unknown X. *Since this latter system reads*

$$\begin{pmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & -5 & -6 \\ 0 & 0 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 6 \\ -23 \\ -3 \end{pmatrix}, \tag{11}$$

We easily get the solution, which is ${}^t(2, 0, 1, 3)$. *Thus, both methods (that are inversion of matrix and LU factorization) provided us with the same solution.*

Remark 2.6.10. *When it comes to solving a square linear systems* $AX = B$ *for several different matrices* B*, the permuted* LU *factorization can be repeatedly applied to solve the equation multiple times for different* B*. In this case it is faster (and more convenient) to do a permuted* LU *decomposition of the matrix* A *once and then solve the triangular matrices system for the different* B*, rather than using Gaussian elimination. Indeed, inverse matrices are very convenient in analytical manipulations, because they allow you to move matrices from one side to the other of equations easily. However, inverse matrices are almost never computed in "serious" numerical calculations. Whenever you see* $A^{-1}B$*, when you go to implement it on a computer you should read* $A^{-1}B$ *as "solve* $AX = B$ *by some method": e.g. solve it by first computing the* LU *factorization of* A *and then using it to solve* $AX = B$.

One reason that you don't usually compute inverse matrices is that it is wasteful: once you have $PA = LU$*, you can solve* $AX = B$ *directly without bothering finding* A^{-1}*, and computing* A^{-1} *requires much more work if you only have to solve a few right-hand sides. Another reason is that for many special matrices, there are ways to solve* $AX = B$ *much more quickly than you can find* A^{-1}*. For example, many large matrices in practice are sparse (mostly zero), and often for sparse matrices you can arrange for* L *and* U *to be sparse too. Sparse matrices are much more efficient to work with than general "dense" matrices because you don't have to multiply (or even store) the zeros. Even if* A *is sparse, however,* A^{-1} *is usually non-sparse, so you lose the special efficiency of sparsity if you compute the inverse matrix.*

The matrices L *and* U *could be thought to have "encoded" the Gaussian elimination process.*

Moreover, The cost of solving a system of linear equations is approximately $\frac{2}{3}p^3$ *floating-point operations if the matrix* A *belngs to* $\mathscr{M}_p(\mathbf{R})$*. This makes it twice as fast as algorithms based on* QR *decomposition, which costs about* $\frac{4}{3}p^3$ *floating-point operations when Householder reflections are used. For this reason,* LU *decomposition is usually preferred.*

The next theorem summarizes the ways of solving a square linear system.

Theorem 2.6.11 (Solving **Square** Linear Systems)**.** *Let A be a matrix in $\mathscr{M}_p(\mathbf{K})$ and B be a column matrix with p rows. Let us consider the following linear system*

$$(\mathcal{S}) \qquad AX = B,$$

where we denoted $X := {}^t(x_1, x_2, \cdots, x_p)$ the unknown vector. Depending on the rank of the matrix A, there are different possible issues when one wants to solve $\mathcal{S}$.

1. 1^{st} *possibility:* $\mathrm{rk}(A) = p$

In this case A is invertible.

- *One can solve $(\mathcal{S})$ by invoquing Point* **5.** *of Theorem 2.6.6. The unique solution of the system $AX = B$ is $((A^{-1}B)_1, (A^{-1}B)_2, \cdots, (A^{-1}B)_p)$. One just have to compute the inverse of A.*
- *One can also solve $(\mathcal{S})$ using the $PA = LU$ method (see Theorem 2.6.8 and the methodology given right below it).*

2. 2^{st} *possibility:* $\mathrm{rk}(A) < p$

In this case A is not invertible.
One can thus use either Gaussian elimination or the permuted LU factorization to determine whether the system has or not a solution. However, the permuted LU factorization is faster and less expensive (in terms of computations) than Gaussian elimination. Note that both these methods still allows one to described the set of solutions precisely.

Let us finish this chapter with an example of linear system, the solution of which is not unique.

Example 2.6.12. *We want to solve the following linear system.*

$$(\mathscr{S}) : \begin{cases} x - z = 1 \\ 2y + 3z = 0 \\ x - z = 1 \end{cases} .$$

To this end, define

$$A := \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 3 \\ 1 & 0 & -1 \end{pmatrix} \qquad B := \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \qquad \& \qquad X := \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

It is clear that $(\mathscr{S})$ is equivalent to $AX = B$. Since the last row of A if full of zeros, it is clear that A is not invertible (since its rank is smaller than p). Our only possibility to solve $(\mathscr{S})$ is thus to use the permuted LU factorization.
We hence start with:

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 3 \\ 1 & 0 & -1 \end{pmatrix} := A_1 \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} := L_1 \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} := P_1$$

The second and last step is therefore:

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 3 \\ 0 & 0 & 0 \end{pmatrix} := A_2 \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} := L_2 \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} := P_2$$

We thus define

$$U := A_2, \qquad L := L_2, \qquad \& \qquad P := P_2.$$

We then have to solve, first:

$$LC = PB \qquad \text{i.e.} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

The solution of such a system is, obviously, $C = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.

We then have to solve:

$$UX = C \qquad \text{i.e.} \qquad \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

Such a system also reads

$$(\mathscr{S}') \begin{cases} x_1 - x_3 = 1 \\ 2x_2 + 3x_3 = 0 \end{cases}$$

$(\mathscr{S}')$ *is equivalent to*

$$\begin{cases} x_1 = 1 + x_3 \\ x_2 = \frac{-3x_3}{2} \end{cases}$$

We see easily see that $(\mathscr{S}')$ *(and thus* $(\mathscr{S})$*) has infinitely many solutions. More precisely, if* $\mathscr{S}ol_{(\mathscr{S})}$ *denotes the set of solutions of* $(\mathscr{S})$*, one can write:*

$$\mathscr{S}ol_{(\mathscr{S})} := \{(1+\alpha, \tfrac{-3\alpha}{2}, \alpha),\ \alpha \in \mathbf{R}\}.$$

Part II
Vector Spaces & Linear Maps

Chapter 3

Vector Spaces

3.1 Vector Space

In the whole chapter $\mathbf{K}$ denotes $\mathbf{R}$ or $\mathbf{C}$.

Definition & Properties

Definition 3.1.1. *Let E be a set endowed with two laws, denoted $+$ and $\cdot$ such that:*

$$\begin{array}{ccc} \mathbf{K} \times E & \to & E \\ (\alpha, \boldsymbol{x}) & \mapsto & \alpha \cdot \boldsymbol{x} \end{array}$$

$(E, +, \cdot)$ is said to be a Vector Space (or a Linear Space) on $\mathbf{K}$ or a $\mathbf{K}$-Vector Space (or a $\mathbf{K}$-Linear Space) if:

- *$(E, +)$ is a commutative (or abelian) group i.e. :*
  1. $\forall (\boldsymbol{x}, \boldsymbol{y}) \in E^2, \quad \boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$
  2. $\forall (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in E^3, \quad (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z} = \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z})$
  3. $\exists\, 0_E \in E$ *s.t.* $\forall \boldsymbol{x} \in E,\ \boldsymbol{x} + 0_E = \boldsymbol{x}$, 0_E *is called the identity element of* $(E, +)$.
  4. $\forall \boldsymbol{x} \in E, \exists - \boldsymbol{x} \in E \quad \boldsymbol{x} + (-\boldsymbol{x}) = 0_E$.
- $\forall (\boldsymbol{x}, \boldsymbol{y}) \in E^2$ *and* $(\alpha, \beta) \in \mathbf{K}^2$, *we have:*
  1. $\alpha \cdot (\boldsymbol{x} + \boldsymbol{y}) = \alpha \cdot \boldsymbol{x} + \alpha \cdot \boldsymbol{y}$
  2. $(\alpha + \beta) \cdot \boldsymbol{x} = \alpha \cdot \boldsymbol{x} + \beta \cdot \boldsymbol{x}$
  3. $1 \cdot \boldsymbol{x} = \boldsymbol{x}$
  4. $\alpha \cdot (\beta \cdot \boldsymbol{x}) = (\alpha\beta) \cdot \boldsymbol{x}$.

  We call ***vectors*** *the elements of E and* ***scalars*** *the elements of $\mathbf{K}$.*

For sake of notational simplicity one will write $x - y$ instead of $x + (-y)$. Starting from these axioms we easily deduce the following properties:

Proposition 3.1.2. *1.* $\forall \boldsymbol{x} \in E,\ 0 \cdot \boldsymbol{x} = 0_E$,

2. $\forall \lambda \in \mathbf{K},\ \lambda \cdot 0_E = 0_E$,

3. $\forall \boldsymbol{x} \in E,\ (-1) \cdot \boldsymbol{x} = -\boldsymbol{x}$,

4. *The following implication holds:* $\alpha \cdot \boldsymbol{x} = 0_E \Longrightarrow \alpha = 0_{\mathbf{K}}$ *or* $\boldsymbol{x} = 0_E$.

Proof. To be written. □

Remark 3.1.3. *1. When there is no ambiguity on* $\mathbf{K}$ *one says Vector Space instead of* $\mathbf{K}$*-Vector Space.*

2. We will often write 0 *instead of* 0_E*. The same symbol will hence denotes, from one hand the neutral element for addition in* E *(and in this case* 0 *belongs to* E*) and, in an other hand, the neutral element for addition in* $\mathbf{K}$ *(and in this case* 0 *belongs to* $\mathbf{K}$*).*

Examples

1. $(\mathbf{R}, +, \cdot)$ and $(\mathbf{C}, +, \cdot)$ are both an $\mathbf{R}$-Vector Space

More generally, The body $(\mathbf{K}, +, \cdot)$ itself, where $+$ denotes the addition between two elements of $\mathbf{K}$ and $\cdot$ denotes the ordinary multiplication between two elements of $\mathbf{K}$ is a $\mathbf{K}$-Vector Space. Indeed, $\cdot$ denotes the ordinary multiplication between two elements of $\mathbf{K}$, *i.e.* :

$$\begin{array}{ccc} \mathbf{K} \times \mathbf{K} & \to & \mathbf{K} \\ (\alpha, x) & \mapsto & \alpha \cdot x := \alpha x \end{array}$$

is a $\mathbf{K}$-Vector Space since $(\mathbf{K}, +, \cdot)$ fulfills all the requirements of a Vector Space on $\mathbf{K}$.

2. $(\mathbf{K}^n, +, \cdot)$ is an $\mathbf{K}$-Vector Space for any n in $\mathbf{N}^*$

Let n be a positive integer. Define $+$ and $\cdot$ on $\mathbf{K}^n$ by setting:

$$(x_1, x_2, \cdots, x_n) + (x'_1, x'_2, \cdots, x'_n) := (x_1 + x'_1, x_2 + x'_2 +, \cdots, x_n + x'_n)$$
$$\alpha \cdot (x_1, x_2, \cdots, x_n) := (\alpha x_1, \alpha x_2, \cdots, \alpha x_n)$$

Let X be a set and let $\mathscr{F}(X, \mathbf{K})$ be the set of $\mathbf{K}$-valued functions, defined on X. Then $(\mathscr{F}(X, \mathbf{K}), +, \cdot)$ is a $\mathbf{K}$-Vector Space if the two laws $+$ and $\cdot$ are defined by setting:

$$\begin{array}{cccc} f+g: & X & \to & \mathbf{K} \\ & t & \mapsto & f(t)+g(t) \end{array} \qquad \begin{array}{cccc} \alpha \cdot f: & X & \to & \mathbf{K} \\ & t & \mapsto & \alpha f(t) \end{array}$$

3. $(\{\boldsymbol{u} := (u_n)_{n \in \mathbf{N}},\ u_n \in \mathbf{K}\}, +, \cdot)$ is a $\mathbf{K}$-Vector Space

The set of all sequences of real or complex numbers, endowed with the addition between sequences and with the multiplication of a sequence by a scalar is Vector Space on $\mathbf{K}$.

4. $(\mathbf{K}[X], +, \cdot)$ is a $\mathbf{K}$-Vector Space.

The set of polynomials with coefficients in $\mathbf{K}$ is Vector Space on $\mathbf{K}$.

5. $(\mathscr{M}_{p,q}(K), +, \cdot)$ is a K-Vector Space.

The set of matrices with p rows and q columns, endowed with the addition between two matrices and with the multiplication by a scalar is Vector Space on $\mathbf{K}$.

Remark 3.1.4. *When there is no risk of confusion we say that E is a vector space instead of $(E, +, \cdot)$.*

3.2 Vector Subspaces

In all this section, $(E, +, \cdot)$ denotes a Vector Space on $\mathbf{K}$.

Definition & Examples

Definition 3.2.1 (Vector subpace). *Let F be a subset of E. $(F, +, \cdot)$ is said to be a vector subspace of $(E, +, \cdot)$ if:*

1. $0_E \in F$,
2. $\forall (x, y) \in F^2, \quad x + y \in F$,
3. $\forall \lambda \in \mathbf{K}, \; \forall x \in F, \; \lambda \cdot x \in F$,

Remark 3.2.2. *It is clear that a Vector subspace is a vector space (for the laws $+$ and $\cdot$ defined on E).*

Example 3.2.3.
1. *$\mathbf{R}$ and $i\mathbf{R}$ are vector subspaces of the $\mathbf{R}$-vector space $\mathbf{C}$.*
2. *E and 0 are vector subspaces of E that are call trivial subspaces of E.*
3. *When F is a sub vector space of E then $(\mathscr{F}(F, \mathbf{K}), +, \cdot)$ is vector subspace of $(\mathscr{F}(E, \mathbf{K}), +, \cdot)$.*
4. *General theorems on continuity show that the set of continuous $\mathbf{R}$-valued functions defined on $[0, 1]$ (denoted $C^0([0, 1], \mathbf{R})$) is a vector subspace of $(\mathscr{F}([0, 1], \mathbf{R}), +, \cdot)$*
5. *The set $F := \{\boldsymbol{x} := (x_1, x_2, \cdots, x_n) \in \mathbf{R}^n, \; \sum_{i=1}^{n} x_i = 0\}$ is a vector subspace of $\mathbf{R}^n$.*
6. *The set $F := \{\boldsymbol{x} := (x_1, x_2, \cdots, x_n) \in \mathbf{R}^n, \; \sum_{i=1}^{n} x_i = 1\}$ is not a vector subspace of $\mathbf{R}^n$ since it does not contain $0_{\mathbf{R}^n}$.*
7. *$\mathbf{K}_n[X] := \{P \in \mathbf{K}[X], \; \deg(P) \leqslant n\}$ is a vector subspace of $\mathbf{K}[X]$.*

3.3 Intersection and Sum of Vector Subspaces

Intersection of Vector subspaces

Proposition 3.3.1. *An intersection of Vector subspaces of E is a vector subspace of E.*

Proof. To be written. □

Proposition 3.3.2. *The union of two vectors subspaces of E that are not included in each other is not a vector subspace of E.*

Proof. To be written. [CPY96, p.48] □

Remark 3.3.3. *One can summarize the previous proposition by writing that an union of Vector subspaces is not, in general, a vector subspace. For example the union of two lines in* $\mathbf{R}^2$ *is not a vector subspace of* $\mathbf{R}^2$.

Definition 3.3.4 (Linear Combination). *Let E be a $\mathbf{K}$-Vector Space and let $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ be vectors of E. We call* ***linear combination*** *of the vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$, any vector of the form:*

$$\sum_{i=1}^{n} \lambda_i\, \boldsymbol{x}_i, \text{ where } n \in \mathbf{N}^*, \text{ and } (\lambda_1, \lambda_2, \cdots, \lambda_n) \in \mathbf{K}^n.$$

If A is a subset of E, we call ***linear combination*** *of elements of A, any linear combination of a finite number of elements of A.*

Remark 3.3.5. *One might want to use sometimes the notation*

$$\sum_{a \in A} \lambda_a\, a.$$

This is not ambiguous if A is a finite subset of E but this suggests, if A is an in a finite subset of E, that only finitely many scalars λ_a are different from 0. Hence the given sum is correctly defined.

Definition 3.3.6 (Collinear vectors). *Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of E are said to be collinear if there exist scalars λ and μ, non both equal to 0, such that*

$$\lambda \boldsymbol{x} + \mu \boldsymbol{y} = 0.$$

If $\boldsymbol{x} \neq 0$, one can not have $\mu \neq 0$ and the condition becomes: there exists a scalar ν such that $\boldsymbol{y} = \nu \boldsymbol{x}$.

If $A := (a_i)_{i \in I}$ is a family of vectors of E, a linear combination of vectors of A is a sum of the form $\sum_{i \in I} \lambda_i\, a_i$ in which all the λ_i equal 0, except finitely many of them. When we index a subset A by its own elements (*i.e.* $I = A$), we recover the notation introduced at remark 3.3.5.

Definition 3.3.7 (Span of A). *We note $\operatorname{Span}(A)$ the set of all linear combinations of elements of A.*

Example 3.3.8. *When $A = \{\boldsymbol{a}\}$ has a unique element, $\operatorname{Span}(A)$ is the set $\mathbf{K}\boldsymbol{a} = \{\lambda \boldsymbol{a},\ \lambda \in \mathbf{K}\}$ of "multiples" of the vector $\boldsymbol{a}$. A set $E := \mathbf{K}\boldsymbol{a}$, where $\boldsymbol{a} \neq 0$ is called a* ***line*** *of E.*

The next proposition shows, in particular, that the lines of E are vector subspaces. This can be checked directly.

Proposition 3.3.9. *Let A be a subset of E (or a family of vectors of E). The set $\operatorname{Span}(A)$ is a vector subspace of E. Moreover, it is the smallest vector subspace which contains A. It is also the intersection of all vector subspaces of E which contain A. It is called the vector subspace spanned by A and we denote it $\operatorname{Span}_{\mathbf{K}}(A)$ or $\operatorname{Span}_{\mathbf{K}} A$ (or only $\operatorname{Span}(A)$ or $\operatorname{Span} A$ when there is no risk of confusion).*

Proof. To be written [CPY96, p.49].
Finally, one easily deduces, from Propositions 3.3.1 and 3.3.9 that $\mathrm{Span}_{\mathbf{K}}(A)$ is the intersection of all vector subspaces of E which contain A. □

Example 3.3.10. *1. If $A = \emptyset$, then $\mathrm{Span}(A) = \{0_E\}$.*

2. In the $\mathbf{R}$ vector space $\mathbf{C}$,

- $\mathrm{Span}_{\mathbf{R}}(\{1\}) = \mathbf{R}$,
- $\mathrm{Span}_{\mathbf{R}}(\{i\}) = i\mathbf{R}$,
- $\mathrm{Span}_{\mathbf{R}}(\{1, i\}) = \mathbf{C}$.

3. Every two non collinear vectors of the plan span $\mathbf{R}^2$.

4. Every three non coplanar vectors of the space span $\mathbf{R}^3$.

5. $(1, X, X^2, \cdots, X^n)$ is a spanning family of $\mathbf{K}_n[X]$ since any polynomial with degree smaller than n can be written under the form:

$$\sum_{k=0}^{n} a_k X^k \text{ with } \forall k \in [\![0, n]\!],\ a_k \in \mathbf{K}.$$

We will see in Proposition 3.4.1 an easier way to prove Statement **5**. of the previous example.

Sum of Vector subspaces

We saw in Remark 3.3.3 that an union of Vector subspaces is not, in general, a vector subspace. The smallest vector subspace of a vector space E which contains F and G is therefore the vector subspace spanned by $F \cup G$ which is, in general different from $F \cup G$.

Proposition 3.3.11. *Let F and G be two vector subspaces of a vector space E. The smallest vector subspace which contains F and G is:*

$$H := \{x + y,\ (x, y) \in F \times G\}.$$

We call it the sum of F and G and denote it $F + G$.

Proof. To be filled! □

Example 3.3.12. *1. If $E = \mathbf{K}^2$ then $E = \mathrm{Span}\{(1, 0)\} + \mathrm{Span}\{(1, 1)\}$ since any element (x, y) of E can be written under the form $(x - y)(1, 0) + y(1, 1)$.*

2. Let F, G and H be three sub-vector spaces of a vector space E. The smallest vector subspace which contains F, G and H is:

$$F + G + H := \{x + y + z,\ (x, y, z) \in F \times G \times H\}.$$

We call it the sum of the vector subspaces F, G and H.

3. More generally, if $E_1, E_2, \cdots, E_n$ are vector subspaces of E, one defines:

$$E_1 + E_2 + \cdots + E_n := \{x_1 + x_2 + \cdots + x_n,\ \ \text{where } \forall i \in [\![1, n]\!],\ x_i \in E_i\}.$$

It is the smallest vector subspace of E which contains every E_i.

Hence we have, for example,

$$\mathrm{Span}\,(a_1, a_2, \cdots, a_n) = \mathbf{K}a_1 + \mathbf{K}a_2 + \cdots + \mathbf{K}a_n.$$

Definition 3.3.13 (Direct Sum of Vector Subspaces). *The sum of two vectors subspaces F and G is said to be direct if $F \cap G = \{0_E\}$. In this case the sum $F + G$ is denoted $F \bigoplus G$.*

Proposition 3.3.14. *Let F and G be two vector subspaces of a vector space E that are in direct sum. For all vector x of $F \bigoplus G$ there exists a unique vector y in F and a unique vector z in G such that $x = y + z$.*

Proof. To be written! [CPY96, p.50]. □

Supplementary Vector Subspaces

Definition 3.3.15 (Direct Sum). *Let F and G be two vector subspaces of a vector space E. We say that F and G are supplementary vector subspaces of E if $E = F \bigoplus G$. In other words, F and G are supplementary vector subspaces of E if $E = F + G$ and if $F \cap G = \{0_E\}$.*
In this latter case, every vector x of E can be written, in a unique way, under the form $x = y + z$ with y in F and z in G.
*Vector y is called the projection of x, on F, parallel to G and the map $f : E \mapsto F$ defined by $f(y+z) = y$, for all (y, z) in $F \times G$, is called the **projection** of E on F, parallel to G.*

Remark 3.3.16. *When F and G are vector subspaces of a vector space E, to prove that $F \cap G = \{0\}$, it is sufficient to prove that $x \in F \cap G \Longrightarrow x = 0$.*

Example 3.3.17. *1. In $E := \mathbf{K}^2$*

- *The two vector subspaces $F := \mathrm{Span}\{(1,0)\}$ and $G := \mathrm{Span}\{(0,1)\}$ are supplementary vector subspaces of $\mathbf{K}^2$. Indeed, any element (x, y) in $\mathbf{K}^2$ can be written uniquely under the form $\alpha(1,0) + \beta(0,1)$, with $\alpha = x$ and $\beta = y$.*
- *The two vector subspaces $F := \mathrm{Span}\{(1,0)\}$ and $H := \mathrm{Span}\{(1,1)\}$ are supplementary vector subspaces of $\mathbf{K}^2$. Indeed,*
  - $\forall (x, y) \in \mathbf{K}^2,\ (x, y) = (x - y) \cdot (1, 0) + y \cdot (1, 1) = (x - y, 0) + (y, y),$
  - *If $(x, 0) = (y, y)$, then $y = 0$ and thus $x = 0$.*

2. In $\mathbf{K}^3$, the vector subspaces $\mathbf{K}^2 \times \{0\}$ and $\mathrm{Span}\{(0,0,1)\}$ are supplementary vector subspaces.

3. In $E := \mathscr{F}(\mathbf{R}, \mathbf{R})$, one can show that the set $\mathscr{E}$ of even functions and the set $\mathscr{O}$ are supplementary vector subspaces (thus $\mathscr{E} \bigoplus \mathscr{O} = \mathscr{F}(\mathbf{R}, \mathbf{R})$).

4. In $E := \mathbf{K}[X]$, let P be in $E \backslash \{0\}$. The sets

$$F := \{Q \in E,\ \deg(Q) < \deg(P)\} \qquad G := \{P \cdot Q,\ Q \in E\}$$

are two supplementary vector subspaces. Use Euclidean division to prove it!

One can generalize Definition 3.3.15 to the case of any finite number of vector subspaces of E.

Definition 3.3.18 (Sum of finitely many vector subspaces). *Let $E_1, E_2, \cdots E_n$ be n vector subspaces of E. Define*

$$E_1 + E_2 + \cdots + E_n = \mathrm{Span}(E_1 \cup E_2 \cup \cdots \cup E_n).$$

By induction, one can prove that

Proposition 3.3.19. *$E_1 + E_2 + \cdots + E_n$ is the set of all vectors of E which can be written under the form of a sum $\boldsymbol{x}_1 + \boldsymbol{x}_2 + \cdots \boldsymbol{x}_n$, where $\boldsymbol{x}_i$ belongs to E_i, for all i in $[\![1, n]\!]$.*

Proof. To be written! [CPY96, p.50]. □

Example 3.3.20. *A family of vectors $(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)$ spans E if and only if $E = \mathbf{K}\boldsymbol{x}_1 + \mathbf{K}\boldsymbol{x}_2 + \cdots + \mathbf{K}\boldsymbol{x}_n$.*

Definition 3.3.21 (Direct Sum). *Let $E_1, E_2, \cdots E_n$ be n vector subspaces of E. We will say that the sum of these vector subspaces is direct if the expansion described in Proposition 3.3.19 is unique for all vector x of E. In this latter case we denote $E_1 \bigoplus E_2 \bigoplus \cdots \bigoplus E_n$ the sum of the E_i.*

Definition 3.3.22 (). *Let E be a vector space. We call **plan** of E every vector subspace which is the direct sum of two lines of E. We call **hyperplan** of E every vector subspace of E which is the supplementary space of a line in E.*

3.4 Spanning Family & Linearly Independent Family of Vectors

We saw at Example 3.3.8 an explicit characterization of $\mathrm{Span}(A)$ when $A = \{\boldsymbol{a}\}$. One can generalize this characterization when A is finite family of vectors of E.

Proposition 3.4.1. *Let n be in $\mathbf{N}^*$ and $A := \{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n\}$ a finite subset of E which contains n elements. The vector subspace spanned by A is the set of all linear combinations of vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n$; i.e.*

$$\mathrm{Span}(A) = \left\{ y \in E,\ \exists(\lambda_1, \lambda_2, \cdots, \lambda_n) \in \mathbf{K}^n \text{ such that } y = \sum_{i=1}^{n} \lambda_i \boldsymbol{a}_i \right\}.$$

In particular, the family $(\boldsymbol{a_1}, \boldsymbol{a_2}, \cdots, \boldsymbol{a_n})$ spans E if

$$E = \mathrm{Span}\,\{\boldsymbol{a_1}, \boldsymbol{a_2}, \cdots, \boldsymbol{a_n}\}$$

i.e. if every element of E can be written as a linear combination of elements of the family $(\boldsymbol{a_1}, \boldsymbol{a_2}, \cdots, \boldsymbol{a_n})$. In other words:

$$\forall x \in E,\ \exists(\lambda_1, \lambda_2, \cdots, \lambda_n) \in \mathbf{K}^n \ \text{ such that } \ x = \sum_{i=1}^{n} \lambda_i \boldsymbol{a_i}.$$

Proof. Obvious in view of Definitions 3.3.4 and 3.3.7. □

Remark 3.4.2.
1. *If a family of vectors spans E, any family of vectors obtained by permuting its elements also spans E. It does not depend on the order of the vectors which belong to the spanning family.*
2. *Note that proving Statement **5.** of Example 3.3.10 is now obvious.*

3. *When the family of vectors one considers in infinite one can not use Proposition 3.4.1 anymore. One has to come back to Definition 3.3.7.*

Example 3.4.3. *A family of vectors* $(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)$ *spans* E *if and only if* $E = \mathbf{K}\boldsymbol{x}_1 + \mathbf{K}\boldsymbol{x}_2 + \cdots + \mathbf{K}\boldsymbol{x}_n$.

Proposition 3.4.4. *Let* $\mathscr{G}$ *be a family of vectors which spans a vector space* E. *A family* $\mathscr{X}$ *of elements of* E *spans* E *if, and only if, every element of* $\mathscr{G}$ *is a linear combination of elements of* $\mathscr{X}$.

Proof. To be written. □

Proposition 3.4.5. *Let* F *and* G *be two vector subspaces of* E. *A vector* x *belongs to* $\operatorname{Span}(F \cup G)$ *if and only if there exists a vector* y *of* F *and a vector* z *of* G *such that* $x = y + z$.

Proof. To be written [CPY96, p.50]. □

The next definition derives naturally from Definition 3.3.4.

Definition 3.4.6 (Linear Independence). *Let* $(\boldsymbol{x}_i)_{i \in [\![1,n]\!]}$ *be a finite family of elements of* E.

1. *We say that* $(\boldsymbol{x}_i)_{i \in [\![1,n]\!]}$ *is a linearly independent family of vectors of* E *if, for all family* $(\lambda_i)_{i \in [\![1,n]\!]}$ *in* $\mathbf{K}^n$, *the following implication holds:*

$$\sum_{i=1}^{n} \lambda_i\, \boldsymbol{x}_i = 0 \Longrightarrow \lambda_1 = \lambda_2 = \cdots \lambda_n = 0.$$

2. *In case the previous implication does not hold, i.e. if one can find a family* $(\lambda_i)_{i \in [\![1,n]\!]}$ *in* $\mathbf{K}^n$ *such that not all the* λ_i *equal* 0, *and such that* $\sum_{i=1}^{n} \lambda_i \boldsymbol{x}_i = 0$, *one says that the family* $(\boldsymbol{x}_i)_{i \in [\![1,n]\!]}$ *is linearly dependent or not linearly independent.*

Remark 3.4.7. *If a family of vectors is linearly independent, any subset of vectors of this family obtained by permuting its elements is also linearly independent. The linear independence of a family of vectors does not depend on the order of its elements.*

Example 3.4.8.

1. *A family with only one element, denoted* $\boldsymbol{x}$, *is linearly independent if and only if* $\boldsymbol{x} \neq 0$. *Indeed,*
   - *If* $\boldsymbol{x} = 0$, *then* $1 \cdot \boldsymbol{x} = 0$ *and* $1 \neq 0$, *and thus the family* $(\boldsymbol{x})$ *is not linearly independent.*
   - *If* $\boldsymbol{x} \neq 0$, *then* $\lambda \cdot \boldsymbol{x} = 0 \Longrightarrow \lambda = 0$.

2. *In the* $\mathbf{R}$*-vector space* $\mathbf{C}$, *the family* $(1, i)$ *is linearly independent since:*

$$\forall (a, b) \in \mathbf{R}^2, a + ib = 0 \Longrightarrow a = b = 0.$$

3. *In* $\mathbf{K}^3$ *vectors* $\boldsymbol{e}_1 := (1, 0, 0)$, $\boldsymbol{e}_2 := (0, 1, 0)$ *and* $\boldsymbol{e}_3 := (0, 0, 1)$ *form a linearly independent family of vectors. Indeed,*

$$\forall (\lambda_1, \lambda_2, \lambda_3) \in \mathbf{K}^3,\ \lambda_1 \boldsymbol{e}_1 + \lambda_2 \boldsymbol{e}_2 + \lambda_3 \boldsymbol{e}_3 = 0 \Longrightarrow (\lambda_1, \lambda_2, \lambda_3) = (0, 0, 0) \Longrightarrow \lambda_1 = \lambda_2 = \lambda_3 = 0.$$

4. *More generally, in* $\mathbf{K}^n$*, the vectors* $\boldsymbol{e}_1 := (1, 0, \cdots, 0), \boldsymbol{e}_2 := (0, 1, 0, \cdots, 0), \cdots, \boldsymbol{e}_n := (0, \cdots, 0, 1)$ *form a linearly independent family of vectors since:*

$$\begin{aligned}\forall(\lambda_1, \cdots, \lambda_n) \in \mathbf{K}^n,\ \lambda_1\boldsymbol{e}_1 + \cdots + \lambda_n\boldsymbol{e}_n = 0 &\Longrightarrow (\lambda_1, \lambda_2, \cdots, \lambda_n) = (0, 0, \cdots, 0)\\ &\Longrightarrow \lambda_1 = \lambda_2 = \cdots = \lambda_n = 0.\end{aligned}$$

5. *The family* $(1, X, X^2, \cdots, X^n)$ *is linearly independent in* $\mathbf{K}[X]$ *since the polynomial* $\sum_{k=0}^{n} \lambda_k X^k$ *equals* 0 *if and only if all* λ_k *equal* 0.

6. *The empty family, i.e. with no vector in it, is linearly independent since there is no linear combination with all non zero coefficient which equals* 0.

Proposition 3.4.9. 1. *Every subset of vectors of a linearly independent family of vectors is linearly independent.*

2. *Every set of vectors containing a non linearly independent family of vectors is not linearly independent.*

Proof. To be written [CPY96, p.50]. □

Remark 3.4.10. 1. *A linearly independent family of vectors can not contain the null vector.*

2. *A linearly independent family can not have two proportional vectors.*

Proposition 3.4.11. *Let* $(\boldsymbol{x}_i)_{i\in[\![1,n]\!]}$ *be a linearly independent family of vectors of* E*. The family of vectors* $(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n, \boldsymbol{x})$ *is linearly dependent if and only if* $\boldsymbol{x}$ *is a linear combination of* $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$.

Proof. To be written. □

Theorem 3.4.12. *Let* $(\boldsymbol{x}_i)_{i\in[\![1,n]\!]}$ *be a family of linearly independent vectors of* E *and let* $(\lambda_1, \cdots, \lambda_n)$ *and* $(\mu_1, \cdots, \mu_n)$ *be two family of scalars, we have the following implication:*

$$\sum_{i=1}^{n} \lambda_i x_i = \sum_{i=1}^{n} \mu_i x_i \Longrightarrow \forall i \in \{1, 2, \cdots, n\},\ \lambda_i = \mu_i.$$

Proof. To be written. □

3.5 Linear Maps

Before beginning this section, readers should be familiar with the notions of injectivity, surjectivity and bijectivity of functions. See Appendix VII if needed.

Let E, F and G be three vector spaces on $\mathbf{K}$.

Definition and characterization

Definition 3.5.1 (Linear Map). *An application $u : E \to F$ is said to be a linear map (also called a linear mapping, linear transformation or linear function) if it preserves the two laws of a vector space:*

$$\forall(\boldsymbol{x},\boldsymbol{y}) \in E^2, \qquad u(\boldsymbol{x}+\boldsymbol{y}) = u(\boldsymbol{x}) + u(\boldsymbol{y}) \qquad \& \qquad u(\lambda\boldsymbol{x}) = \lambda u(\boldsymbol{x}) \tag{1}$$

These two equalities can be gathered under the form:

$$\forall(\boldsymbol{x},\boldsymbol{y}) \in E^2, \qquad u(\lambda\boldsymbol{x}+\boldsymbol{y}) = \lambda u(\boldsymbol{x}) + u(\boldsymbol{y}) \tag{2}$$

We also say that u is a morphism between vector spaces E and F. We say that u is an:

- *endomorphism if $E = F$,*
- *isomorphism if u is bijective,*
- *automorphism if u is bijective and if $E = F$.*

Notations The set of linear maps from E to F is denoted $\mathscr{L}(E,F)$ or $\mathscr{L}_{\mathbf{K}}(E,F)$ if one wants to precise the body $\mathbf{K}$. Moreover, $\mathscr{L}(E) := \mathscr{L}(E,E)$ denotes the set of endomorphisms of E.

Remark 3.5.2. *If $u : E \to F$ is a linear map then*

1. *we have $u(0_E) = 0_F$ since:*
$$u(0_E) = u(0_{\mathbf{R}} \cdot 0_E) = 0_F.$$
2. $$\forall \boldsymbol{x} \in E,\ \ u(-\boldsymbol{x}) = -u(\boldsymbol{x}).$$
3. *We immediately can prove (by induction on n in $\mathbf{N}$) that, for all $(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)$ in E^n and $(\lambda_1, \lambda_2, \cdots, \lambda_n)$ in $\mathbf{R}^n$:*
$$u(\sum_{i=1}^{n} \lambda_i\, \boldsymbol{x}_i) = \sum_{i=1}^{n} \lambda_i\, u(x_i).$$

Example 3.5.3.
1. *The linear maps from $\mathbf{R}$ to $\mathbf{R}$ are the maps $x \mapsto kx$, with k in $\mathbf{R}$. (Try to prove it).*
2. *For any (a,b) in $\mathbf{R}^2$, the map $u : \begin{array}{ccc} \mathbf{R}^2 & \to & \mathbf{R} \\ (\boldsymbol{x},\boldsymbol{y}) & \mapsto & ax+by \end{array}$ is linear.*
3. *For any (a,b,c,d) in $\mathbf{R}^4$, the map $u : \begin{array}{ccc} \mathbf{R}^2 & \to & \mathbf{R}^2 \\ (x,y) & \mapsto & (ax+by, cx+dy) \end{array}$ is an endomorphism of $\mathbf{R}^2$.*
4. *The differentiation is a linear map from $\mathscr{C}^1(\mathbf{R})$ in $\mathscr{C}^0(\mathbf{R})$. This means that:*
$$\psi : \begin{array}{ccc} \mathscr{C}^1(\mathbf{R}) & \to & \mathscr{C}^0(\mathbf{R}) \\ f & \mapsto & f' \end{array}$$
Note in particular that the restriction of ψ to $\mathscr{C}^\infty(\mathbf{R})$ is an endomorphism of $\mathscr{C}^\infty(\mathbf{R})$.
5. *For all (a,b,c) in $\mathbf{R}^3$, the map:*
$$\Psi : \begin{array}{ccc} \mathscr{C}^2(\mathbf{R}) & \to & \mathscr{C}^0(\mathbf{R}) \\ f & \mapsto & af''+bf'+cf \end{array}$$
is a linear map.

6. *If F and G are supplementary vector subspaces of E, one can check that the map, from E to F, "projection on F, parallel to G" is linear.*

Definition 3.5.4 (Linear form). *We call linear form any linear map from E to $\mathbf{R}$.*

Example 3.5.5.

1. *For any (α, β) in $\mathbf{R}^2$, the map* $f : \begin{array}{ccc} \mathbf{R}^2 & \to & \mathbf{R} \\ (x, y) & \mapsto & ax + by \end{array}$ *is a linear form on $\mathbf{R}^2$.*

2. *On the segment $[a, b]$, the Riemann Integral is a linear form on $\mathscr{C}^0(I)$. Indeed, it is easy to verify that:*

$$\Lambda : \begin{array}{ccc} \mathscr{C}^0(I) & \to & \mathbf{R} \text{ is linear.} \\ f & \mapsto & \int_a^b f(t)\, dt \end{array}$$

is linear.

3. *For any α in $\mathbf{R}$, the map $P \mapsto P(\alpha)$ is a linear form on $\mathbf{R}[X]$.*

Structure of $\mathscr{L}(E, F)$ and $\mathscr{L}(E)$.

Proposition 3.5.6. *Let f and g be two endomorphisms of E and α and β be two scalars. Then the map $\alpha f + \beta g$ is linear.*

Proposition 3.5.7.

1. *If $f \in \mathscr{L}(E, F)$ and $g \in \mathscr{L}(F, G)$ then $g \circ f$ belongs to $\mathscr{L}(E, G)$.*
2. *If f is an isomorphism from the vector spaces E and F, then f^{-1} is linear.*
3. *$(\mathscr{L}(E, F), +, \cdot)$ is a vector space.*

Proof. To be written! [CPY96, Proposition II.17 p.52]. □

Proposition 3.5.8. *Let $u \in \mathscr{L}(E,F)$ and $v \in \mathscr{L}(F,G)$ and (α,β) in $\mathbf{R}^2$.*

1. *Let φ and ψ be two linear applications from F to G, we have:*

$$(\alpha\varphi + \beta\psi) \circ u = \alpha(\varphi \circ u) + \beta(\psi \circ u).$$

2. *Let φ and ψ be two linear applications from E to F, we have:*

$$v \circ (\alpha\varphi + \beta\psi) = \alpha(v \circ \varphi) + \beta(v \circ \psi).$$

In other words, the maps:

$$\begin{array}{ccc} \mathscr{L}(F,G) & \to & \mathscr{L}(E,G) \\ \varphi & \mapsto & \varphi \circ u \end{array} \quad \text{and} \quad \begin{array}{ccc} \mathscr{L}(E,F) & \to & \mathscr{L}(E,G) \\ \varphi & \mapsto & v \circ \varphi \end{array}$$

are linear.

3. *$(\mathscr{L}(E), +, \circ)$ is a **non commutative** ring[a] in general.*

[a] Give the definition of ring!!!!

Definition 3.5.9 (Linear group). *We call $GL(E)$ the set of automorphisms of the vector space E.*

Proposition 3.5.10. *$(GL(E), \circ)$ is a group. We call it the Linear Group of E.*

Kernel and Image of a linear map

Recall

For any sets E and F and any map $u : E \to F$, we call:

- The image by u of a subset E' of E the set:

$$u(E') := \{u(\boldsymbol{x}),\ \boldsymbol{x} \in E'\},$$

- The preimage by u of a subset F' of F the set:

$$u^{-1}(F') := \{\boldsymbol{x} \in E,\ u(\boldsymbol{x}) \in F'\}.$$

Proposition 3.5.11. *Let $f : E \to F$ be a linear map.*

- *If E' is vector subspace of E, then $f(E')$ is a vector subspace of F.*
- *If F' is vector subspace of F, then $f^{-1}(F')$ is a vector subspace of E.*

Definition 3.5.12 (). *Let $u : E \to F$ be a linear map. We call:*

1. *Kernel of u (or null space of u) the set defined by:*

$$\mathrm{Ker}(u) := u^{-1}(\{0_F\}) = \{x \in E,\ u(\boldsymbol{x}) = 0_F\},$$

2. *Image of u the set defined by:*

$$\mathrm{Im}(u) := u(E) = \{u(\boldsymbol{x}),\ \boldsymbol{x} \in E\},$$

Remark 3.5.13. *Note that one can also denote $N(f)$ the kernel of f.*

Proposition 3.5.14. *Let $u : E \to F$ be a linear map. Then the set $\mathrm{Ker}(u)$ is a vector subspaces of E and the set $\mathrm{Im}(u)$ is a vector subspaces of F.*

Proof. To be written! [CPY96, Proposition II.19 p.52]. □

Theorem 3.5.15. *Let $u : E \to F$ be a linear map. The following statements are equivalent.*

1. *u is injective.*
2. *$Ker(u) = \{0_E\}$.*
3. *$\forall x \in E,\ u(x) = 0 \Longrightarrow x = 0_E$.*

Proof. To be written! [CPY96, Proposition II.20 p.52]. □

Remark 3.5.16.
1. *When one knows that a map is linear, in order to show that it is injective one usually uses Property 2. or its translation 3.*
2. *Let $u : E \to E'$ be a linear map.*
   - *If F and G are two vector subspaces of E, one has:*

   $$u(F + G) = u(F) + u(G).$$

   - *More generally, if $E_1, E_2, \cdots, E_n$ are vector subspaces of E then*

   $$u(E_1 + E_2 + \cdot + E_n) = u(E_1) + u(E_2) + \cdots + u(E_n).$$

3.6 Basis

Definition 3.6.1 (Basis of a Vector Space). *A family of elements of E is a basis of E if it is a linearly independent family which spans E.*

The following example gives basis of some classical vector spaces. The linear independence of the family of vectors presented below has been established in Example 3.4.8, on page 38.

Example 3.6.2. *1. $(1, i)$ constitutes a basis of the $\mathbf{R}$-vector space $\mathbf{C}$.*

2. *The family constituted with vectors $\boldsymbol{e}_1 := (1,0,0)$, $\boldsymbol{e}_2 := (0,1,0)$ and $\boldsymbol{e}_3 := (0,0,1)$ forms a basis of $\mathbf{K}^3$. Note that $\mathscr{B} := (\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$ is called* ***the standard basis of*** $\mathbf{K}^3$. *Moreover, any vector $\boldsymbol{x} := (x_1, x_2, x_3)$ of $\mathbf{K}^3$ can be written as $\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + x_3\boldsymbol{e}_3$. Thus $(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$*

3. *More generally, we call* ***standard basis of*** $\mathbf{K}^n$ *the family of vectors $(\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n)$ where, for every k in $[\![1, n]\!]$, $e_k := (0, 0, \cdots, 0, 1, 0, \cdots, 0)$ i.e. all entries are 0 except for the k^{th} which equals 1.*

4. *The family $(1, X, X^2, \cdots, X^n)$ is a basis of $\mathbf{K}_n[X]$ since every single polynomial $\sum_{k=0}^{n} \lambda_k\ X^k$ equals 0 if and only if all λ_k equal 0. it is called the* ***standard basis of*** $\mathbf{K}_n[X]$.

5. *The family $(X^k)_{k\in\mathbf{N}}$ is a basis of the vector space of polynomials.*

6. *Let p and q be two positive integers. Denote $E_{i,j}$ the $p \times q$ matrix the coefficients of which are all zeros except for the one on the i^{th} row and j^{th} column which equals 1. Let $A := (a_{i,j})_{(i,j)\in[\![1,p]\!]\times[\![1,q]\!]}$ be a matrix in $\mathscr{M}_{p,q}(\mathbf{K})$. We clearly have the equality:*

$$A = \sum_{i=1}^{p}\sum_{i=1}^{q} a_{ij}E_{i,j}.$$

One then deduces that the family $(E_{i,j})_{(i,j)\in[\![1,p]\!]\times[\![1,q]\!]}$ is a basis of $\mathscr{M}_{p,q}(\mathbf{K})$.

Remark 3.6.3. *The two following formula might be of interest for some exercises we will perform later.*

$${}^{t}E_{i,j} = E_{i,j} \qquad \& \qquad E_{i,j}\cdot E_{k,l} = \begin{cases} (0) & \text{if } j \neq k \\ E_{i,l} & \text{if } j = k \end{cases}$$

Theorem 3.6.4. *A family of vectors $(\boldsymbol{e}_i)_{i\in[\![1,n]\!]}$ is a basis of E if, and only if, for all vectors $\boldsymbol{x}$ in E, there exists a* ***unique*** *family $(\boldsymbol{\lambda}_i)_{i\in[\![1,n]\!]}$ of $\mathbf{K}^n$ such that:*

$$\boldsymbol{x} = \sum_{i=1}^{n} \lambda_i\ \boldsymbol{e}_i.$$

The family $(\lambda_i)_{i\in[\![1,n]\!]}$ is called family of components or coordinates of $\boldsymbol{x}$ in the basis $\mathscr{B} := \{\boldsymbol{e}_i,\ i \in [\![1, n]\!]\}$.

Proof. To be filled! □

Remark 3.6.5. *For any given family family of vectors $(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)$ of E, one can easily prove that the map:*

$$\begin{array}{ccc} \mathbf{K}^n & \to & E \\ (\lambda_i)_{i\in[\![1,n]\!]} & \mapsto & \sum_{i=1}^{n} \lambda_i\ \boldsymbol{x}_i \end{array}$$

is linear. Moreover,

- *it is injective (or one to one) if, the family $(\boldsymbol{x}_i)_{i\in[\![1,n]\!]}$ is linearly independent.*
- *it is surjective (or onto) if, the family $(\boldsymbol{x}_i)_{i\in[\![1,n]\!]}$ spans E.*
- *it is bijective if, the family $(\boldsymbol{x}_i)_{i\in[\![1,n]\!]}$ is a basis of E.*

Basis and Linear Maps

Proposition 3.6.6. *Let $u : E \to F$ be a linear map and let $A := (\boldsymbol{a}_i)_{i\in I}$ be a family of vectors of E.*

1. *If u is injective then the family A is linearly independent in E if and only if $(u(\boldsymbol{a}_i))_{i\in I}$ is linearly independent in F.*
2. *If u is surjective then the family A spans E if and only if $(u(\boldsymbol{a}_i))_{i\in I}$ spans F.*
3. *If u is bijective then the family A is a basis of E if and only if $(u(\boldsymbol{a}_i))_{i\in I}$ is a basis of F.*

Proof. To be written [CPY96, Proposition II.33 p.56]. □

Proposition 3.6.7. *Let E and F be two vector spaces. Let $A := (\boldsymbol{a}_i)_{i\in I}$ be a basis of vectors of E and $B := (\boldsymbol{b}_i)_{i\in J}$ be a family of vectors of F. There exists a unique linear map u from E to F such that:*

$$\forall i \in I,\ \ u(\boldsymbol{a}_i) = \boldsymbol{b}_i.$$

Moreover,

u is injective if, and only if, B is linearly independent.

u is surjective if, and only if, B spans F.

u is bijective if, and only if, B is a basis of F.

3.7 Linear Equations

Definition & Examples

To be filled! p.809

Structure of the Set of solutions

To be filled! p.810

Chapter 4

Vector Spaces with finite Dimension

In all this chapter, n is a positive integer and E denotes z vector space on $\mathbf{C}$. moreover $\mathbf{K}$ denotes the set $\mathbf{R}$ of real numbers.

Definition 4.0.1 (Finite Dimension). *A Vector space E is said to be finite-dimensional, if it admits a finite spanning family. Otherwise, we say that E is infinite-dimensional.*

4.1 First Results

Proposition 4.1.1. *In a vector space E which has a spaning family with n elements, every family that has $n+1$ elements is not linearly independent.*

Proof. To be written [CPY96, Proposition II.37 p.57]. ☐

Theorem 4.1.2 (Definition of the dimension)**.** *A vector space with finite dimension has basis. They all have the same (finite) number of elements. This number is called the* ***dimension*** *of the vector space E and is denoted $\dim E$ (or $\dim_{\mathbf{K}} E$ if one wants to specify the body $\mathbf{K}$).*

Proof. To be written [CPY96, Proposition II.38 p.58]. ☐

Proposition 4.1.3. *Let E be a vector space with finite dimension n.*

1. *Every linearly independent family of vectors of E has at most n elements.*
2. *Every spanning family of vectors of E has at least n elements.*
3. *Every linearly independent family of vectors of E, which has n elements is a basis of E.*

Proof. To be written [CPY96, Proposition II.39 p.58]. ☐

Remark 4.1.4. *According to the previous proposition, if $(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_p)$ is a linearly independent family and if $(\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_q)$ is spanning family of E, then $p \leqslant q$.*
Thus every spanning family has, at least, as many elements as any linearly independent family. We therefore deduce the two following results.

Theorem 4.1.5. *Let E be a finite-dimensional vector space, all basis of E have the same number of elements, denoted n. The positive integer n is called* ***dimension of*** E ***on*** $\mathbf{K}$*, or more simply* ***dimension*** *of E.*

Proof. To be filled! □

Example 4.1.6.

1. *The vector space $\{0\}$ has a dimension equal to 0.*
2. $\mathbf{K}$ *has a dimension equal to 1 on $\mathbf{K}$.*
3. $\mathbf{C}$ *has a dimension equal to 2 on $\mathbf{R}$, since $(1, i)$ is a basis of $\mathbf{C}$.*
4. $\mathbf{K}^n$ *is a vector space with dimension equals to n, since a basis of it is constituted by:*

$$\boldsymbol{e}_1 := (1, 0, \cdots, 0), \qquad \boldsymbol{e}_2 := (0, 1, \cdots, 0), \qquad \cdots, \qquad \boldsymbol{e}_n := (0, \cdots, 0, 1),$$

5. $\mathbf{K}_n[X]$ *is a vector space with dimension equals to $n+1$ on $\mathbf{K}$, one basis of $\mathbf{K}_n[X]$ is $(1, X, X^2, \cdots, X^n)$.*

Corollary 4.1.7. *Let E be a finite dimensional vector space of dimension n and let $(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_p)$ a family with p elements of E.*

- *If $p > n$, the family $(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_p)$ is not linearly independent.*
- *If $p < n$, the family $(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_p)$ does not span E.*

Example 4.1.8. *To be filled!!!*

4.2 Dimension of Vector Subspaces

Theorem 4.2.1 (Dimension of the Vector Subspaces)**.** *Let E be a finite-dimensional vector space. Let F be a vector subspace of E. Then the vector space F is finite dimensional. Moreover we have the the relation $\dim F \leqslant \dim E$. Moreover F equals E if, and only if, $\dim F = \dim E$.*

Proof. To be written [CPY96, Theorem II.40 p.59]. □

Proposition 4.2.2. *A vector space E is infinite-dimensional if, and only if, there exists a sequence $(x_n)_{n\in\mathbf{N}}$ of elements of E such that for all n in $\mathbf{N}$, the family of vectors $(x_k)_{0\leqslant k\leqslant n}$ is linearly independent.*

Proof. To be filled! □

Example 4.2.3. *The vector space $K[X]$ is infinite-dimensional.*

Complete a Family of Vectors into a Basis

Theorem 4.2.4 (Completion of a family of vectors)**.** *Let E be a finite-dimensional vector space with $n := \dim E$. Let A be a family of vectors which spans E and let $B := (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_p)$ be a linearly independent family of vectors of E. There exists vectors, denoted $\boldsymbol{a}_{p+1}, \cdots, \boldsymbol{a}_n$ of A such that the family of vectors $(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_p, \boldsymbol{a}_{p+1} \cdots, \boldsymbol{a}_n)$ is a basis of E.*

Proof. To be written [CPY96, Theorem II.42 p.59]. □

Corollary 4.2.5. *Let E be a finite-dimensional vector space with $n := \dim E$.*

1. *From every spanning family A of E one can extract a basis of E.*
2. *Every linearly independent family of E can be completed so that it becomes a basis of E (in general there are infinitely many ways to do so).*
3. *Every spanning family of E with n elements is a basis of E.*

Proof. To be filled! See [CPY96, p.59 Corolaire II.43] □

Remark 4.2.6. *If $E = \{0\}$, the only linearly independent family is the empty family which, in this case, the only basis of E.*

Theorem 4.2.7. *Let $\mathscr{B}$ be a family of elements of a vector space E of finite dimension n. The following statements are equivalent:*

(i) *$\mathscr{B}$ is a basis of E,*

(ii) *$\mathscr{B}$ is a linearly independent family of E with n elements.*

(iii) *$\mathscr{B}$ is a family of n vectors which spans E.*

Proof. To be filled! □

Remark 4.2.8. *This fundamental result is of a great use to show that a family $\mathscr{B}$ is a basis of a vector space the dimension of which is known. In most cases we prove that $\mathscr{B}$ is a linearly independent family with n elements.*

Example 4.2.9.

1. *In the $\mathbf{R}$ vector space $\mathbf{C}$ the family $(1, \omega)$ is linearly independent, i.e. is a basis, if and only if, ω is not a real number.*
2. *Two non-colinear vectors of a vector space of dimension 2 form a basis.*

Example 4.2.10. ***Linear recurrence relation of order*** 2
To be filled!

Example 4.2.11.

1. *In $E = \mathbf{R}^4$, denote the vector space*

$$F := \{(x_1, x_2, x_3, x_4),\ x_1 + x_2 + x_3 + x_4 = 0\}.$$

A vector $\boldsymbol{x} := (x_1, x_2, x_3, x_4)$ of F is characterized by its first three elements x_1, x_2 and x_3. This leads us to think that $\dim(F) = 3$. One can effectively prove that $\dim(F) = 3$ by giving the three following vectors of F:

$$\boldsymbol{f}_1 := (1, 0, 0, -1), \qquad \boldsymbol{f}_2 := (0, 1, 0, -1), \qquad \boldsymbol{f}_3 := (0, 0, 1, -1).$$

Let us prove that $\mathscr{F} := (\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$ is a basis of F by showing that:

(a) *$\mathscr{F}$ is a linearly independent family of vectors. Let $(\lambda_1, \lambda_2, \lambda_3)$ be in $\mathbf{R}^3$.*

$$\begin{aligned}\lambda_1 \boldsymbol{f}_1 + \lambda_2 \boldsymbol{f}_2 + \lambda_3 \boldsymbol{f}_3 = 0 &\Longrightarrow (\lambda_1, \lambda_2, \lambda_3, -\lambda_1 - \lambda_2 - \lambda_3) = 0 \\ &\Longrightarrow \lambda_1 = \lambda_2 = \lambda_3 = 0.\end{aligned}$$

This last implication establishes that $\mathscr{F}$ is a linearly independent family of vectors.

(b) *$\mathscr{F}$ spans F. Let $\boldsymbol{x} := (x_1, x_2, x_3, x_4)$ be an element of F. By definition of F, on has the equality $\boldsymbol{x} = (x_1, x_2, x_3, -x_1 - x_2 - x_3)$. Thus one can write $\boldsymbol{x} = x_1 \boldsymbol{f}_1 + x_2 \boldsymbol{f}_2 + x_3 \boldsymbol{f}_3$. This latter equality, as well as the arbitrary choice of $\boldsymbol{x}$ in F proves that $\mathscr{F}$ spans F.*

2. *In $E = \mathbf{R}^4$, denote the vector space*

$$G := \{(x_1, x_2, x_3, x_4),\ x_1 + x_2 = 0 \ \text{and}\ x_3 + x_4 = 0\}.$$

Any vector $\boldsymbol{x} := (x_1, x_2, x_3, x_4)$ of G fulfills the equality $\boldsymbol{x} = (x_1, -x_1, -x_4, x_4)$. This leads us to think that $\dim(G) = 2$. Let's prove it by showing that the family of vectors $\mathscr{G} := (\boldsymbol{g}_1, \boldsymbol{g}_2)$ defined by by setting:

$$\boldsymbol{g}_1 := (1, -1, 0, 0), \qquad \boldsymbol{g}_2 := (0, 0, -1, 1)$$

constitutes a basis of G.

(a) *$\mathscr{G}$ is a linearly independent family of vectors. Let (λ_1, λ_2) be in $\mathbf{R}^2$.*

$$\begin{aligned}\lambda_1 \boldsymbol{g}_1 + \lambda_2 \boldsymbol{g}_2 = 0 &\Longrightarrow (\lambda_1, -\lambda_1, -\lambda_4, \lambda_4) = 0 \\ &\Longrightarrow \lambda_1 = \lambda_4 = 0.\end{aligned}$$

This last implication establishes that $\mathscr{G}$ is a linearly independent family of vectors.

(b) *$\mathscr{G}$ spans G. Let $\boldsymbol{x} := (x_1, x_2, x_3, x_4)$ be an element of G. By definition of G, on has the equality $\boldsymbol{x} := (x_1, -x_1, -x_4, x_4)$. Thus one can write $\boldsymbol{x} = x_1 \boldsymbol{g}_1 + x_4 \boldsymbol{g}_2$. This latter equality, as well as the arbitrary choice of $\boldsymbol{x}$ in G proves that $\mathscr{G}$ spans G.*

3. *Since $\mathbf{K}$ is a 1-dimensional vector space on $\mathbf{K}$, its vector subspaces can only have a dimension equal to 0 or 1. Thus $\mathbf{K}$ has onmy to vector subspaces that are $\{0\}$ and $\mathbf{K}$.*

Proposition 4.2.12. *Any vector subspace F of a finite dimensional vector space E has, at least one supplementary vector space G. Thus $F \bigoplus G = E$.*

Proof. To be filled! See [CPY96, p.59 Proposition II.44] □

Remark 4.2.13. *Unless F equals $\{0_E\}$ or E, a vector subspace has many supplementary vector subspaces. It is therefore very important to speak about **a** supplementary vector subspace.*

4.3 Link Between Dimensions

Dimension and Isomorphism

Proposition 4.3.1. *Let E be a finite-dimensional $\mathbf{K}$-vector space of dimension n. A vector space F is isomorphic to E if and only if F is a finite-dimensional vector space of dimension n.*

Proof. To be filled! ☐

Corollary 4.3.2. *Any $\mathbf{K}$-vector space of dimension n is isomorphic to $\mathbf{K}^n$.*

Proof. To be filled! ☐

Dimension of a Product of Vector Spaces

Proposition 4.3.3. *Let E and F be two finite dimensional vector spaces, then $E \times F$ is a finite dimensional vector space, the dimension of which is*

$$\dim(E \times F) = \dim(E) \cdot \dim(F).$$

Proof. To be filled! ☐

Example 4.3.4. *Donner l'exemple de $\mathscr{M}_{p,q}(\mathbf{R}) \simeq \mathbf{R}^p \times \mathbf{R}^q$!*

Dimension of a Sum of Vector Spaces

Proposition 4.3.5. *Let E be a $\mathbf{K}$-vector Space of finite dimension. Let F and G be two vector subspaces of E, we have the following equality:*

$$\dim(F + G) = \dim(F) + \dim(G) - \dim(F \cap G).$$

Proof. To be filled! See [CPY96, p.60 Proposition II.45] ☐

Example 4.3.6. *Expand [CPY96, p.62 Example II.46]*

4.4 Rank Nullity Theorem

Rank of Family of Vectors, of a Linear Map

Definition 4.4.1 (Rank). *The rank of a finite family $\mathscr{X}$ of vectors of E, denoted $\operatorname{rk}(\mathscr{X})$ or $\operatorname{rk}\mathscr{X}$, is the dimension of the vector subspace of E spanned by $\mathscr{X}$.*

Remark 4.4.2. *Let $\mathscr{X} := (x_1, x_2, \cdots, x_n)$ be a family of vectors of E.*

1. *Since $\mathscr{X}$ is a spanning family of $\operatorname{Span}\mathscr{X}$, one can extract a basis of $\operatorname{Span}\mathscr{X}$. Thus $\operatorname{rk}\mathscr{X} \leqslant n$ and the equality holds if and only if $\mathscr{X}$ is a linearly independent family.*

2. *If $\mathscr{X}$ has r linearly independent vectors, we have $\operatorname{rk}\mathscr{X} \leqslant r$, and the equality case happens if and only if these r linearly independent vectors span $\operatorname{Span}\mathscr{X}$, in other words the equality happens if and only if any element of $\mathscr{X}$ is a linear combination of these r vectors.*

Definition 4.4.3 (Rank of a linear map). *Let E and F be two $\mathbf{K}$ vector spaces and $u : E \to F$ be a linear map. We call rank of u, and we denote it $\operatorname{rk}(u)$, the dimension of $\operatorname{Im}(u)$, when it is finite.*

Remark 4.4.4.

- *If F is a finite-dimensional space with dimension n, the rank of any linear map $u : E \to F$ is finite and smaller or equal to n. moreover, the equality holds if and only if u is surjective.*
- *Let E be a finite-dimensional vector space with dimension p and $\mathscr{B} := (\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_p)$ a basis of E. If $u : E \to F$ is a linear map, we saw at Proposition ?? that $\operatorname{Im}(u)$ is spanned by the family of vectors $(u(\boldsymbol{e}_k))_{1\leqslant k\leqslant n}$. This proves that $\operatorname{rk}(u)$ is finite. It is equal to $\operatorname{rk}(u(\mathscr{B}))$. Thus is smaller or equal to p and is equal , if and only if, $u(\mathscr{B})$ is a basis of $\operatorname{Im}(u)$ (i.e. if and only if u is injective).*

Rank Nullity Theorem

Theorem 4.4.5 (Rank Nullity Theorem)**.** *Let E be a finite-dimensional vector space and a $u : E \to F$ a linear map valued in a vector space F. Then $\operatorname{Im}(u)$is a finite-dimensional vector space. moreover one has the following equality:*

$$\dim(E) = \dim(\operatorname{Ker}(u)) + \dim(\operatorname{Im}(u)). \tag{1}$$

Remark 4.4.6. *In view of Definition 5.2.6, the previous equality can be rewritten under the form*

$$\dim(E) = \dim(\operatorname{Ker}(u)) + \operatorname{rk}(u). \tag{2}$$

Example 4.4.7.

1. *Let $h : \mathbf{R}^2 \to \mathbf{R}^2$ be the linear map defined, for all (x, y) in $\mathbf{R}^2$, by setting:*

$$h(x, y) := (x - y, -3x + 3y).$$

It is clear that $\operatorname{Ker}(h) = \operatorname{Span}\{(1, 1)\}$. According to the rank nullity theorem we therefore know that $\dim(\operatorname{Im}(h)) = 2 - 1 = 1$. Thus $\operatorname{Im}(h) = \operatorname{Span}\{h((1, 0))\} = \operatorname{Span}\{(1, -1)\}$.

2. *Let Δ be the endomorphism of $\mathbf{R}_n[X]$ defined by:*

$$\Delta(P) = P(X + 1) - P(X).$$

- *Every constant polynomial is in the kernel of Δ. Conversely, if $\Delta(P) = 0$, then $P(X + 1) = P(X)$ and thus all integers are roots of the polynomial $P(X) - P(0)$. It is therefore null and thus P is constant. This implies that $\operatorname{Ker}\Delta = \mathbf{R}_0[X]$.*
- *According to the rank nullity theorem, $\dim(\operatorname{Im}(\Delta)) = n$. Since $\operatorname{Im}(\Delta) \subset \mathbf{R}_{n-1}[X]$, the dimension of which is also n, we deduce that $\operatorname{Im}\Delta = \mathbf{R}_{n-1}[X]$.*

From these two points one can deduce that every polynomials with a degree smaller than n is of the form $P(X + 1) - P(X)$, where $\deg(P) \leqslant n$

Corollary 4.4.8. *Two finite-dimensional vector spaces E and F have the same dimension if and only if there exist a bijective linear map $u : E \to F$.*

Proof. To be filled! See [CPY96, p.62 Corollaire II.48] ☐

Remark 4.4.9. *Let $u : E \to F$ be a bijective linear map. Its inverse function, denoted u^{-1}, is also linear.*

Linear Maps and Finite Dimension

Corollary 4.4.10 (Characterization of Isomorphisms). *Let E and F be two finite-dimensional vector spaces with the same dimension, denoted n and let $u : E \to F$ be a linear map. The following properties are equivalent;*

(i) *u is injective,*

(ii) *u is surjective,*

(iii) *u is bijective.*

Proof. To be filled! See [CPY96, p.62 Corollaire II.50] $\square$

Example 4.4.11. *To be written!*

Corollary 4.4.12. *If u is an endomorphism, of a finite-dimensional vector space E, it is equivalent to to say that the map u is injective, or surjective or bijective.*

Proof. Evidence absolue au vu du corollaire précédent. $\square$

Remark 4.4.13. *Be careful if E is not a finite-dimensional vector space!*

- *An endomorphism of E may be injective without being surjective, as the example of the following linear map shows:*

$$\begin{array}{rccl} \Phi : & \mathbf{R}[X] & \to & \mathbf{R}[X] \\ & P & \mapsto & X \cdot P(X). \end{array}$$

Indeed, Φ is clearly linear. Moreover, $\mathrm{Ker}(\Phi) = \{P \in \mathbf{R}[X],\ X \cdot P(X) = 0\} = \{P := \sum_{k=0}^{q} a_k X^k, q \in \mathbf{N}, (a_0, \cdots, a_q) \in \mathbf{R}^{q+1}$ s.t. $\sum_{k=0}^{q} a_k X^{k+1} = 0\}$. Thus $\mathrm{Ker}(\Phi) = \{0\}$ and therefore Φ is injective. However $P(X) := 1$ does not belong to $\mathrm{Im}(\Phi)$ which is therefore not surjective.

- *An endomorphism of E may be surjective without being injective, as the example of the following linear map shows:*

$$\begin{array}{rccl} \Psi : & \mathbf{R}[X] & \to & \mathbf{R}[X] \\ & P & \mapsto & P'(X). \end{array}$$

Indeed, any constant polynomial has an image by Ψ which equals 0 so Ψ is not injective. Besides any polynomial $P := \sum_{k=0}^{q} a_k X^k$ of $\mathbf{R}[X]$ fulfills the equality

$$\Psi\left(\sum_{k=0}^{q} \frac{a_k}{k+1} X^{k+1}\right) = P.$$

This makes Ψ surjective.

4.5 The four fundamental subspaces

To be filled (See Strang the 4 pages in Ressources/ books Linear Algebra)

Chapter 5

Use of Basis

In this chapter, we show how the language of Vector Spaces can be translated into Matrix language and vice-versa. More precisely, we are considering finite-dimensional vector spaces and we choose a basis for each of them. We first show that the objects defined in Chapters 3 and 4 (vectors spaces and linear maps) have a matrix representation. Then we study how these representations depend on the choice of these basis.

5.1 Matrix Representation

Vectors

Let E *be a finite-dimensional vector space with* $\dim(E) = p$ *and let* $\boldsymbol{e} := (\boldsymbol{e}_1, \cdots, \boldsymbol{e}_p)$ *be a basis of* E. *For every vector* $\boldsymbol{x}$ *of* E, *we have a unique expansion*

$$\boldsymbol{x} = \sum_{i=1}^{p} \alpha_i \boldsymbol{e}_i$$

where the α_i *are scalar (they are the coordinates or components) of vector* $\boldsymbol{x}$ *in the basis* $\mathbf{e}$. *We write:*

$$[\boldsymbol{x}]^e = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix}.$$

In other words, $[\boldsymbol{x}]^e$ *is the column matrix of the coordinates of vector* $\boldsymbol{x}$. *We say that* ***the column matrix represents the vector*** $\boldsymbol{x}$ ***in the basis*** $\mathbf{e}$.
More generally, if $\boldsymbol{x}_1, \cdots \boldsymbol{x}_q$ *are vectors of* E, *we denote* $([\boldsymbol{x}_1]^e \cdots [\boldsymbol{x}_q]^e)$ *or simply* $[\boldsymbol{x}_1, \cdots \boldsymbol{x}_q]^e$ *the* $p \times q$ *matrix the* j^{th} *column of which is* $[\boldsymbol{x}_j]^e$.
We easily verify that the map $\boldsymbol{x} \mapsto [\boldsymbol{x}]^e$, *from* E *to* $\mathscr{M}_{p,1}(\mathbf{K})$, *is linear, i.e. , for all* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *in* E, *we have* $[\lambda \boldsymbol{x} + \boldsymbol{y}]^e = \lambda [\boldsymbol{x}]^e + [\boldsymbol{y}]^e$. *Moreover, since this map transforms the basis* $\mathbf{e}$ *of* E *into the standard basis of* $\mathscr{M}_{p,1}(\mathbf{K})$, *it is bijective (according to Proposition 3.6.7).*

Linear Maps

Let E *be a finite-dimensional vector space with* $\dim(E) = p$ *and let* F *be a finite-dimensional vector space with* $\dim(F) = q$. *Let* $\boldsymbol{e} := (\boldsymbol{e}_1, \cdots, \boldsymbol{e}_p)$ *be a basis of* E, $\boldsymbol{f} := (\boldsymbol{f}_1, \cdots, \boldsymbol{f}_q)$ *be a basis of* F *and* $u : E \to F$ *be a linear map. We denote* λ_{ij} $(1 \leqslant i \leqslant q)$ *the coordinates of the vector* $u(\boldsymbol{e}_j)$ *in the basis* $\boldsymbol{f}$, *i.e. the scalars which verify:*

$$u(\boldsymbol{e}_j) = \sum_{i=1}^{q} \lambda_{ij}\ \boldsymbol{f}_i.$$

We denote $Mat_{e,\boldsymbol{f}}(u)$ *or* $[\boldsymbol{u}]_{\boldsymbol{f}}^{e}$ *the* $p \times q$ *matrix* $(\lambda_{ij})_{1\leqslant i\leqslant q, 1\leqslant j\leqslant p}$. *It is called the matrix of the linear map* u *between the basis* $\mathbf{e}$ *and* $\boldsymbol{f}$.
It is clear that the columns of the matrix $[\boldsymbol{u}]_{\boldsymbol{f}}^{e}$ *are the column matrices* $[\boldsymbol{u}(\boldsymbol{e}_j)]^{\boldsymbol{f}}$. *In other words:*
The matrix of a linear map $u : E \to F$, *between basis* $\mathbf{e}$ *and* $\boldsymbol{f}$, *has, as columns, the column matrices of the coordinates of the image of the vectors of basis* $\mathbf{e}$ *of* E, *in the basis* $\boldsymbol{f}$ *of* F:

$$[\boldsymbol{u}]_{e}^{\boldsymbol{f}} = \Big([\boldsymbol{u}(\boldsymbol{e}_1)]^{\boldsymbol{f}} \cdots [\boldsymbol{u}(\boldsymbol{e}_p)]^{\boldsymbol{f}}\Big)\,.$$

Example 5.1.1. *Let* $u : \mathbf{R}^3 \to \mathbf{R}^2$ *be defined, for all* (x, y, z) *in* $\mathbf{R}^3$, *by setting:*

$$u(x, y, z) := (x + y + z, -x + 2y + 2z).$$

Define $\mathscr{B} := (\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$ *the standard basis of* $\mathbf{R}^3$ *and* $\mathscr{B}' := (\boldsymbol{f}_1, \boldsymbol{f}_2)$ *the standard basis of* $\mathbf{R}^2$. *Denote* A *the matrix of the linear map* u *between the basis* $\mathbf{e}$ *and* $\boldsymbol{f}$. *We easily compute*

$$u(\boldsymbol{e}_1) = (\mathbf{1}, -\mathbf{1}) = \mathbf{1}\boldsymbol{f}_1 - \mathbf{1}\boldsymbol{f}_2 \qquad u(\boldsymbol{e}_2) = (1, 2) \qquad u(\boldsymbol{e}_3) = (1, 2).$$

Thus we can write:

$$A = [\boldsymbol{u}]_{e}^{\boldsymbol{f}} = \begin{array}{cc} \begin{array}{ccc} u(\boldsymbol{e}_1) & u(\boldsymbol{e}_2) & u(\boldsymbol{e}_3) \end{array} & \\ \begin{pmatrix} 1 & 1 & 1 \\ -1 & 2 & 2 \end{pmatrix} & \begin{array}{c} \boldsymbol{f}_1 \\ \boldsymbol{f}_2 \end{array} \end{array}$$

It is clear that, if u *and* v *are both linear maps in* $\mathscr{L}(E, F)$, *we have, for all scalar* λ,

$$[\lambda \boldsymbol{u} + \boldsymbol{v}]_{e}^{\boldsymbol{f}} = \lambda\,[\boldsymbol{u}]_{e}^{\boldsymbol{f}} + [\boldsymbol{v}]_{e}^{\boldsymbol{f}}$$

Remark 5.1.2. *Note that one can also refer to* $[u]_{e}^{\boldsymbol{f}}$ *as* $Mat_{e,\boldsymbol{f}}(u)$.

Fondamental Formulae

We will see in this paragraph how the notations we introduced in the previous paragraph are adapted to the operations already defined, from one hand on Matrices, and, from the other hand on linear maps.
Let E *be a finite-dimensional vector space with* $\dim(E) = p$, F *be a finite-dimensional vector space with* $\dim(F) = q$ *and* $u : E \to F$ *be a linear map. With the notations introduced at 5.1 and at 5.1, we easily see that, for all vector* $\boldsymbol{x}$ *in* E,

$$\begin{aligned} u(\boldsymbol{x}) &= u\left(\sum_{j=1}^{p} \alpha_j\ \boldsymbol{e}_j\right) = \sum_{j=1}^{p} \alpha_j\ u(\boldsymbol{e}_j) \\ &= \sum_{j=1}^{p} \alpha_j \sum_{i=1}^{q} \lambda_{ij}\ \boldsymbol{f}_i = \sum_{i=1}^{q} \left(\sum_{j=1}^{p} \lambda_{ij}\ \alpha_j\right) \boldsymbol{f}_i. \end{aligned}$$

This latter formula can be naturally summarized by:

$$[u(\boldsymbol{x})]^{\boldsymbol{f}} = [\boldsymbol{u}]_{e}^{\boldsymbol{f}} \cdot [\boldsymbol{x}]^{e}\,. \tag{1}$$

Conversely, if A is a matrix in $\mathscr{M}_{p,q}(\mathbf{K})$, the map $u : X \mapsto AX$, from $\mathscr{M}_{p,1}(\mathbf{K})$ to $\mathscr{M}_{q,1}(\mathbf{K})$ is a linear map. Ifone identifies $\mathscr{M}_{p,1}(\mathbf{K})$ with $\mathbf{K}^p$ and $\mathscr{M}_{q,1}(\mathbf{K})$ with $\mathbf{K}^q$ using standard basis $\boldsymbol{e}$ of $\mathbf{K}^p$ and $\boldsymbol{f}$ of $\mathbf{K}^q$, we see that, for all vector X of $\mathbf{K}^p$, we have the equality:

$$[u]_{\boldsymbol{e}}^{\boldsymbol{f}} X = [\boldsymbol{u}]_{\boldsymbol{e}}^{\boldsymbol{f}} [X]^{\boldsymbol{f}} = [u(X)]^{\boldsymbol{f}} = u(X) = AX.$$

We deduce that $[\boldsymbol{u}]_{\boldsymbol{e}}^{\boldsymbol{f}} = A$.
Let E be a finite-dimensional vector space with $\dim(E) = p$, F be a finite-dimensional vector space with $\dim(F) = q$ and G be a finite-dimensional vector space with $\dim(G) = r$. Let $\boldsymbol{e}$, $\boldsymbol{f}$ and $\boldsymbol{g}$ be basis of E, F and G and let $u : E \to F$ and $v : F \to G$ be two linear maps. We see that, for all vector X of E, we have the equality:

$$[v \circ u]_{\boldsymbol{e}}^{\boldsymbol{g}} = [v]_{\boldsymbol{f}}^{\boldsymbol{g}} [u]_{\boldsymbol{e}}^{\boldsymbol{f}}. \tag{2}$$

One can also check directly the previous formula by writting:

$$v \circ u(\boldsymbol{e}_j) = v\left(\sum_{i=1}^{q} \lambda_{ij} \boldsymbol{f}_i\right) = \sum_{i=1}^{q} \lambda_{ij} v(\boldsymbol{f}_i) = \sum_{i=1}^{q} \lambda_{ij} \sum_{k=1}^{r} \mu_{ki}\, \boldsymbol{g}_k = \sum_{i=1}^{q}\left(\sum_{k=1}^{r} \lambda_{ij}\, \mu_{ki}\right) \boldsymbol{g}_k.$$

A Particular Linear Map

Endomorphisms

Endomorphisms
Let E be a finite-dimensional vector space with $\dim(E) = p$, $\boldsymbol{e}$ be a basis of E and u be an endomorphism of E. Since both the departure and the arrival sets are the same, it is natural to take the same basis. The matrix $[\boldsymbol{u}]_{\boldsymbol{e}}^{e}$ is called the matrix of the endomorphism u in the basis $\boldsymbol{e}$.
hence, whatever the basis $\boldsymbol{e}$ of E, we have $Id_E(\boldsymbol{e}_i) = \boldsymbol{e}_i$. Therefore, $[Id_E]_{\boldsymbol{e}}^{e} = I_p$. Besides, Applying Formula (2)*, we get $[v \circ u]_{\boldsymbol{e}}^{e} = [u]_{\boldsymbol{e}}^{e} \cdot [u]_{\boldsymbol{e}}^{e}$. Thus, by induction, we get*

$$[u^n]_{\boldsymbol{e}}^{e} = ([u]_{\boldsymbol{e}}^{e})^n, \quad \forall n \in \mathbf{N}. \tag{3}$$

If u is bijective then this formula holds for every n in $\mathbf{Z}$.

Proposition 5.1.3. *Let E and F be two finite-dimensional vector spaces with $\dim(E) = p$ and $\dim(F) = q$. Let $\boldsymbol{e} := (\boldsymbol{e}_1, \cdots, \boldsymbol{e}_p)$ be a basis of E, $\boldsymbol{f} := (\boldsymbol{f}_1, \cdots, \boldsymbol{f}_p)$ be a basis of F and $u : E \to F$ be a linear map. The map u is bijective if, and only if, $p = q$ and if the matrix $[\boldsymbol{u}]_{\boldsymbol{e}}^{\boldsymbol{f}}$ is invertible. In this latter case, we have the equality: $\left[\boldsymbol{u}^{-1}\right]_{\boldsymbol{f}}^{e} = \left([\boldsymbol{u}]_{\boldsymbol{e}}^{\boldsymbol{f}}\right)^{-1}$.*

Proof. To be filled! See [CPY96, p.81 Proposition III.1] □

Rank of a Linear Map

Theorem 5.1.4. *Let E and F be two finite-dimensional vector spaces, $\boldsymbol{e}$ being a basis of E and $\boldsymbol{f}$ being a basis of F and let $u : E \to F$ be a linear map. The rank of the matrix $[\boldsymbol{u}]_{\boldsymbol{e}}^{\boldsymbol{f}}$ equals the rank of the linear map u. In particular it does not depend on the choice of the basis $\boldsymbol{e}$ nor $\boldsymbol{f}$*

Proof. To be filled! See [CPY96, p.82 Théorème III.2] □

Corollary 5.1.5. *Let q be a non-zero integer, $S := (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_q)$ be a family of vectors of a finite-dimensional vector space E and $\boldsymbol{e}$ be a basis of E. The rank of S (which is by definition the rank of vector subspace $\mathrm{Span}(S)$ equals the rank of the matrix*

$$([\boldsymbol{x}_1]^{\boldsymbol{e}} \cdots [\boldsymbol{x}_q]^{\boldsymbol{e}}) .$$

Proof. To be filled! See [CPY96, p.84 Corollaire III.3] ☐

5.2 Change of Basis

Notations

Let E be a finite-dimensional vector space with $\dim(E) = p$ and let $\boldsymbol{e}$ and $\boldsymbol{f}$ be two basis of E. In order to make a clear disctinction between these two basis, we will use the following terminology.

Change of Basis Matrix
The change of basis matrix, denoted P, between the former basis $\boldsymbol{e}$ and the new one $\boldsymbol{f}$ (or simply the change of basis matrix P from $\boldsymbol{e}$ to $\boldsymbol{f}$) is the matrix the columns of which represent the vectors of the new basis, given in the coordinate system of the former basis.

$$P = [I_E]^{\boldsymbol{e}}_{\boldsymbol{f}} = \left([\boldsymbol{f}_1]^{\boldsymbol{e}} \cdots \left[\boldsymbol{f}_p\right]^{\boldsymbol{e}}\right) . \tag{4}$$

Proposition 5.2.1. *The change of basis matrix P from the basis $\boldsymbol{e}$ to $\boldsymbol{f}$. Its inverse P^{-1} is the change of basis matrix from the basis $\boldsymbol{f}$ to $\boldsymbol{e}$.*

Proof. To be filled! See [CPY96, p.84 Proposition III.4] ☐

Example 5.2.2. *Let $\boldsymbol{e} := (\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$ be the standard basis of $\mathbf{R}^3$ and define*

$$\begin{cases} \boldsymbol{f}_1 := 2\boldsymbol{e}_2 + 3\boldsymbol{e}_3 \\ \boldsymbol{f}_2 := 2\boldsymbol{e}_1 - 5\boldsymbol{e}_2 - 8\boldsymbol{e}_3 \\ \boldsymbol{f}_3 := -\boldsymbol{e}_1 + 4\boldsymbol{e}_2 + 6\boldsymbol{e}_3. \end{cases}$$

Define $\boldsymbol{f} := (\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$. One can easily verify that $\boldsymbol{f}$ is a linearly independent family of vectors and therefore a basis of $\mathbf{R}^3$. Denote P the change of basis matrix from $\boldsymbol{e}$ to $\boldsymbol{f}$. According to (4), we can write that:

$$P := [I_E]^{\boldsymbol{e}}_{\boldsymbol{f}} := \begin{pmatrix} 0 & 2 & -1 \\ 2 & -5 & 4 \\ 3 & -8 & 6 \end{pmatrix},$$

and then compute P^{-1}. One then gets:

$$P^{-1} = [I_E]^{\boldsymbol{f}}_{\boldsymbol{e}} = \begin{pmatrix} 2 & -4 & 3 \\ 0 & 3 & -2 \\ -1 & 6 & -4 \end{pmatrix}.$$

The case of vectors

The components of a vector $\boldsymbol{x}$ are modified by a change of basis (from $\mathbf{e}$ to $\boldsymbol{f}$ let's say). Precisely the relation is:

$$[x]^{\boldsymbol{f}} = [I_E(x)]^{\boldsymbol{f}} = [I_E]_{\boldsymbol{e}}^{\boldsymbol{f}} \cdot [x]^{\boldsymbol{e}} .$$

One can translate this equality by writing:

Let X be the column matrix of the components of a vector in the former basis, Y be the column matrix of the components of the same vector in the new basis and let P denote the change of basis matrix. We then have the following equality:

$$Y = P^{-1}X. \tag{5}$$

Example 5.2.3. *We keep the data given at Example 5.2.11. Denote $\boldsymbol{x}$ the vector defined by $\boldsymbol{x} := 2\boldsymbol{e}_1 - 3\boldsymbol{e}_2 + 4\boldsymbol{e}_3$. Denote also $X := [x]^{e}$ and $Y := [x]^{\boldsymbol{f}}$. It is clear that*

$$[x]^{e} = \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix} .$$

According to (5)*, it is easy to determine the components of the vector $\boldsymbol{x}$ in the basis $\boldsymbol{f}$. We hence can write:*

$$Y = [x]^{\boldsymbol{f}} = [I_E(x)]^{\boldsymbol{f}} = [I_E]_{\boldsymbol{e}}^{\boldsymbol{f}} \cdot [x]^{e} = P^{-1} \cdot X = \begin{pmatrix} 2 & -4 & 3 \\ 0 & 3 & -2 \\ -1 & 6 & -4 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix} = \begin{pmatrix} 28 \\ -17 \\ -36 \end{pmatrix} .$$

The case of Linear Maps

Let E and F be two finite-dimensional vector spaces with $\dim(E) = p$ and $\dim(F) = q$. Let $\mathbf{e}$ and $\mathbf{e}'$ be two basis of E, $\boldsymbol{f}$ and $\boldsymbol{f}'$ be two basis of F and let $u : E \to F$ be a linear map. The obvious equality $u = I_F \circ I_E$ can be translated by:

$$[\boldsymbol{u}]_{\boldsymbol{e}'}^{\boldsymbol{f}'} = [\boldsymbol{I}_F]_{\boldsymbol{f}}^{\boldsymbol{f}'} \cdot [\boldsymbol{u}]_{\boldsymbol{e}}^{\boldsymbol{f}} \cdot [\boldsymbol{I}_E]_{\boldsymbol{e}'}^{\boldsymbol{e}} . \tag{6}$$

In order to remember this formula, we introduce below standard conventions (that must be respected if one does not want to make mistakes). Basis $\mathbf{e}$ and $\boldsymbol{f}$ are said to be "former basis" while $\mathbf{e}'$ and $\boldsymbol{f}'$ are said to be "new basis".

Let A be the matrix of the linear map $u : E \to F$ in the former basis and let B be the matrix of the linear map u in the new basis. Denote P the change of basis matrix in E (i.e. from $\mathbf{e}$ to $\boldsymbol{f}$) and let Q be the change of basis matrix in F (i.e. from $\mathbf{e}'$ to $\boldsymbol{f}'$). We have the following equality:

$$B = Q^{-1}AP. \tag{7}$$

Proposition 5.2.4. *Let A be a square matrix of size p and of rank r. There exist two square invertible matrices, denoted G and D such that $GAD = J_r$, where we defined:*

$$J_r = \begin{pmatrix} 1 & 0 & & & \cdots & 0 \\ 0 & \ddots & & & & \\ & & 1 & & & \\ \vdots & & & 0 & & \vdots \\ & & & & \ddots & \\ 0 & & & \cdots & & 0 \end{pmatrix} \begin{matrix} \\ \\ \leftarrow \textit{row } r \\ \\ \\ \\ \end{matrix}$$

Proof. To be filled! See [CPY96, p.85 Proposition III.5] ☐

Remark 5.2.5. 1. *Note that J_r can also be denoted by*

$$\begin{pmatrix} I_r & 0_{p-r,p-r} \\ 0_{p-r,p-r} & 0_{r,r} \end{pmatrix} \tag{8}$$

where $0_{s,t}$ denotes the null matrix which has s rows and t columns.

2. *Multiply a matrix A, from the right hand side (respectively from the left hand side) by a square and invertible matrix comes to make a change of basis in the departure space (respectively to make a change of basis in the arrival space) of the map u. It does not modify or affect the rank of the matrix: since A and B both represent the linear map u, we have the equality $\mathrm{rk}(A) = \mathrm{rk}\, u = \mathrm{rk}(B)$. Since the rank of J_r is obviously r, we see that the property given in Proposition 6.3.2 characterizes the matrices with rank r.*

3. *The way to obtain matrices G and D is to perform Gaussian Elimination (see chapter 1 for more details).*

Definition 5.2.6 (Rank of a linear map). *Let E and F be two $\mathbf{K}$ vector spaces and $u : E \to F$ be a linear map. We call rank of u, and we denote it $\mathrm{rk}(u)$, the dimension of $\mathrm{Im}(u)$, when it is finite.*

Corollary 5.2.7. *The rank of a matrix equals the rank of its transpose.*

Proof. To be filled! See [CPY96, p.86 Corollaire III.7] ☐

Corollary 5.2.8. *The rank of a $p \times q$ matrix equals the rank of the family of its rows vectors in $\mathbf{K}^q$.*

Proof. To be filled! See [CPY96, p.86 Corollaire III.8] ☐

Definition 5.2.9 (Equivalent Matrices). *Two matrices A and B of $\mathscr{M}_{p,q}(\mathbf{K})$ are said to be equivalent when it exists a $q \times q$ invertible matrix, denoted P, and a $p \times p$ invertible matrix, denoted Q such that:*

$$B = QAP.$$

Remark 5.2.10. *It results from the previous definition that all matrices which represent a given linear map u are equivalent to each other.*

The case of Endomorphisms

In the case where the linear map u is an endomorphism of a vector space E, there is, obviously, only one former basis, and only one new basis. The rule therefore becomes:

> *Let A be the matrix of an endomorphism $u : E \to E$ in the former basis and let B be the matrix of the endomorphism u in the new basis. Denote P the change of basis matrix. We have the following equality:*
>
> $$B = P^{-1}AP. \tag{9}$$

Example 5.2.11. *Define the endomorphism of $\mathbf{R}^3$ by setting:*

$$\begin{cases} u(\boldsymbol{e}_1) := 2\boldsymbol{e}_1 - 3\boldsymbol{e}_2 + 2\boldsymbol{e}_3 \\ u(\boldsymbol{e}_2) := -3\boldsymbol{e}_1 - 6\boldsymbol{e}_2 + 6\boldsymbol{e}_3 \\ u(\boldsymbol{e}_3) := 2\boldsymbol{e}_1 - 2\boldsymbol{e}_2 + 2\boldsymbol{e}_3. \end{cases}$$

It is clear that the we have the equality:

$$M := [u]_{\boldsymbol{e}}^{\boldsymbol{e}} = Mat_{\boldsymbol{e}}(u) = \begin{pmatrix} 2 & -3 & 2 \\ -3 & -6 & -2 \\ 2 & 6 & 2 \end{pmatrix}$$

where $Mat_{\boldsymbol{e}}(u)$ is a mathematical shorthand for $Mat_{\boldsymbol{e},\boldsymbol{e}}(u)$, introduced at Remark 5.1.2. Having in mind the definition of the basis vectors $\boldsymbol{f} := (\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$ of E, given at Example 5.2.2, denote P the change of basis matrix from $\boldsymbol{e}$ to $\boldsymbol{f}$. We already know that

$$P := \begin{pmatrix} 0 & 2 & -1 \\ 2 & -5 & 4 \\ 3 & -8 & 6 \end{pmatrix}$$

and that:

$$P^{-1} = \begin{pmatrix} 2 & -4 & 3 \\ 0 & 3 & -2 \\ -1 & 6 & -4 \end{pmatrix}.$$

We can then use (9) to write that:

$$[\boldsymbol{u}]_{\boldsymbol{f}}^{\boldsymbol{f}} = [\boldsymbol{I}_E]_{\boldsymbol{e}}^{\boldsymbol{f}} \cdot [\boldsymbol{u}]_{\boldsymbol{e}}^{\boldsymbol{e}} \cdot [\boldsymbol{I}_E]_{\boldsymbol{f}}^{\boldsymbol{e}}.$$

In other words we have:

$$\begin{aligned} Mat_{\boldsymbol{f}}(u) = P^{-1} \cdot M \cdot P &= \begin{pmatrix} 2 & -4 & 3 \\ 0 & 3 & -2 \\ -1 & 6 & -4 \end{pmatrix} \cdot \begin{pmatrix} 2 & -3 & 2 \\ -3 & -6 & -2 \\ 2 & 6 & 2 \end{pmatrix} \cdot \begin{pmatrix} 0 & 2 & -1 \\ 2 & -5 & 4 \\ 3 & -8 & 6 \end{pmatrix} \\ &= \begin{pmatrix} 126 & -280 & 230 \\ -90 & 204 & -167 \\ -180 & 405 & -332 \end{pmatrix}. \end{aligned}$$

Definition 5.2.12 (Similar Matrices). *Two matrices A and B of $\mathscr{M}_p(\mathbf{K})$ are said to be similar when there exists a $p \times p$ invertible matrix P such that:*

$$B = P^{-1}AP.$$

Remark 5.2.13. *It results from the previous definition that all matrices which represent the same given endomorphism u are similar to each other.*

Part III

Multilinear Maps, Determinants

Chapter 6

Multilinear Maps & Determinants

Rajouter des exemples pour chaque section car il n'y en n'a pas un seul!

In this entire chapter p denotes a positive integer.

6.1 Definitions

They are many different ways to define the notion of Determinant. Of course they are all equivalent. In this course we will start from p-linear function to do so.
We identify in the sequel a vector $\boldsymbol{x}$ of $\mathbf{K}^p$ with the column matrix $[\boldsymbol{x}]^e$, where $\boldsymbol{e}$ denotes the standard basis of $\mathbf{K}^p$ (in other words one write the components of the vector $\boldsymbol{x}$ in column). In particular it is the same thing to consider a family $(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_p)$ of p vectors of $\mathbf{K}^p$ or to consider the $p \times p$ matrix $A := ([\boldsymbol{x}_1]^e \cdots [\boldsymbol{x}_p]^e)$, the columns of which are the column matrices $[\boldsymbol{x}_j]^e$.

Definition 6.1.1 (Alternating p-linear form). *An Alternating p-linear form on $\mathbf{K}^p$ is a map $\mathfrak{b} : \mathscr{M}_p(\mathbf{K}) \to \mathbf{K}$ such that:*

1. *Whatever the index i, when one freezes all the columns of a matrix $A := ([\boldsymbol{x}_1]^e \cdots [\boldsymbol{x}_i]^e \cdots [\boldsymbol{x}_p]^e)$, except for the i^{th}, one gets a map $X_i \mapsto \mathfrak{b}\left([\boldsymbol{x}_1]^e \cdots [\boldsymbol{x}_{i-1}]^e\, X_i\, [\boldsymbol{x}_{i+1}]^e \cdots [\boldsymbol{x}_p]^e\right)$ from $\mathbf{K}^p$ to $\mathbf{K}$ which is linear.*
2. *If the matrix A has two equal columns ($X_i = X_j$ for two different indices i and j) then $\mathfrak{b}(A) = 0$.*

Proposition 6.1.2. *Let $\mathfrak{b}$ be an alternating p-linear form on $\mathbf{K}^p$ and B be the matrix obtained by exchanging two columns of Matrix A. One has the equality $\mathfrak{b}(B) = -\mathfrak{b}(A)$.*

Proof. To be filled! See [CPY96, p.106 Proposition IV.13] □

Proposition 6.1.3. *An alternating p-linear form on $\mathbf{K}^p$ is completely determined by the value of $\mathfrak{b}(I_p)$, where I_p denotes the identity matrix of size p.*

Proof. To be filled! See [CPY96, p.106 Proposition IV.14] □

Remark 6.1.4. *It is easy to verify that the set $\mathfrak{B}$ of the alternating p-linear forms on $\mathbf{K}^p$ is a vector subspace of the vector space $\mathbf{K}^{\mathscr{M}_p(\mathbf{K})}$. The previous proposition shows that two alternating p-linear forms on $\mathbf{K}^p$ are collinear. In other words, the vector space $\mathfrak{B}$ has a dimension not greater than 1. In order to show that $\dim(\mathfrak{B})$ is 1 exactly, one just needs to find an alternating p-linear forms on $\mathbf{K}^p$ different from 0.*

Let $\mathfrak{b}$ be the alternating p-linear form defined on $\mathscr{M}_p(\mathbf{K})$ by setting

$$\mathfrak{b}(A) := \sum_{\sigma \in \sigma_p} \operatorname{sgn}(\sigma)\, a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(p),p} \cdot \mathfrak{b}(I_p), \tag{1}$$

where $A := (a_{i,j})_{(i,j)\in[\![1,p]\!]}$ belongs to $\mathscr{M}_p(\mathbf{K})$, σ_p denotes the set of all permutations[1] of $\{1, \cdots, p\}$ and $\operatorname{sgn}(\sigma)$ denotes the signature of the permutation σ. In (1), the sum is computed over all permutations σ of the set $\{1, 2, ..., p\}$.

Remark 6.1.5. *Note that* (1) *also reads*

$$\mathfrak{b}(A) := \sum_{\sigma \in \sigma_p} \operatorname{sgn}(\sigma) \prod_{i=1}^{p} a_{\sigma(i),i} \cdot \mathfrak{b}(I_p). \tag{2}$$

One can check that the form $\mathfrak{b}$, defined in (1)*, for which we impose that $\mathfrak{b}(I_p) := 1$, is a non zero alternating linear form. This is this particular linear form we will be interested in in the sequel.*

Definition 6.1.6 (Determinant). *We call determinant, and we denote **det**, the alternating p-linear form on $\mathbf{K}^p$ such that **det**$(I_p) = 1$.*

Remark 6.1.7.

1. *Gathering both Equality* (2) *and Definition 6.1.6 we obtain the Leibniz formula, which reads for an $p \times p$ matrix A:*

$$\mathbf{det}(A) := \sum_{\sigma \in \sigma_p} \operatorname{sgn}(\sigma) \prod_{i=1}^{p} a_{\sigma(i),i}. \tag{3}$$

2. *Note that the determinant is only defined for square matrices.*

3. *We sometimes use the notation*

$$\begin{vmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,q} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,q} \\ \vdots & \vdots & & \vdots \\ a_{p,1} & a_{p,2} & \cdots & a_{p,q} \end{vmatrix}$$

*to denote **det**(A).*

[1] A permutation is a function that reorders this set of integers (here the set is $1, 2, ..., p$). The value in the i^{th} position after the reordering σ is denoted by $\sigma(i)$. For example, for $p = 3$, the original sequence $1, 2, 3$ might be reordered to $\sigma := (2, 3, 1)$ (*i.e.* $\sigma(1) = 2$, $\sigma(2) = 3$ and $\sigma(3) = 1$). The set of all such permutations (also known as the symmetric group on p elements) is denoted by σ_p. For each permutation σ, $\operatorname{sgn}(\sigma)$ denotes the signature of σ, a value that is $+1$ whenever the reordering given by σ can be achieved by successively interchanging two entries an even number of times, and -1 whenever it can be achieved by an odd number of such interchanges.

6.2 Properties of Determinants

Theorem 6.2.1. *Determinant have the following properties.*

1. *A determinant of a matrix with contains two identical columns is zero.*
2. *A determinant of a matrix which has a column full of zeros.*
3. *When one changes two columns of a determinant the sign changes.*
4. *The sign of a determinant will not change if one adds, to one of its columns a linear combination of its other columns.*
5. *If the family of column matrices $A_1, \cdots, A_p$ is not linearly independent (i.e. the corresponding family of vectors of $\mathbf{K}^p$ is not linearly independent) then $\boldsymbol{det}(A_{1,\cdots A_p}) = 0$.*

Proof. To be filled! See [CPY96, p.108 Proposition IV.17] □

A direct consequence of the p-linearity of the determinant is that, for all λ in $\mathbf{K}$ and A in $\mathscr{M}_p(\mathbf{K})$, we have:

$$\boldsymbol{det}(\lambda A) = \lambda^p \boldsymbol{det}(\lambda A). \tag{4}$$

Theorem 6.2.2. *Let A and B be two matrices in $\mathscr{M}_p(\mathbf{K})$. We have the equality:*

$$\boldsymbol{det}(AB) = \boldsymbol{det}(A) \cdot \boldsymbol{det}(B). \tag{5}$$

Proof. To be filled! See [CPY96, p.109 Theorème IV.18] □

Theorem 6.2.3. *In order the $p \times p$ matrix A is invertible it is necessary and sufficient that its determinant is non zero. In this latter case, one has the equality:*

$$\boldsymbol{det}(A^{-1}) = \frac{1}{\boldsymbol{det}(A)}. \tag{6}$$

Proof. To be filled! See [CPY96, p.109 Theorème IV.19] □

Theorem 6.2.4. *For every matrix A of $\mathscr{M}_p(\mathbf{K})$, the following equality holds:*

$$\boldsymbol{det}({}^tA) = \boldsymbol{det}(A).$$

Proof. To be filled! See [CPY96, p.109 Theorème IV.19] □

Theorem 6.2.5. *Determinant have the following properties.*

1. *A determinant of a matrix with contains two identical rows is zero.*
2. *A determinant of a matrix which has a row full of zeros.*
3. *When one exchanges two rows of a determinant the sign changes.*
4. *The sign of a determinant will not change if one adds, to one of its rows a linear combination of its other rows.*
5. *If the rows of matrices of a square matrix A are not linearly independent (i.e. the corresponding family of vectors of $\mathbf{K}^p$ is not linearly independent) then $\boldsymbol{d}et(A) = 0$.*

Proof. One applies Theorem to the matrix ${}^tA$ since the columns of this latter are nothing but the rows of A. ☐

6.3 Expanding a Determinant Along a Row or a Column

Definition 6.3.1 (). *Let $A := (a_{ij})$ be a matrix of $\mathscr{M}_p(\mathbf{K})$. We call minor of the coefficient a_{ij} in A, and we denote it $M_{ij}(A)$, the determinant of the matrix of $\mathscr{M}_{p-1}(\mathbf{K})$ obtained by removing the i^{th} row and the j^{th} column of A.*

Proposition 6.3.2. *We have the equality:*

$$\boldsymbol{d}et(A) = \sum_{j=1}^{p} a_{ij}(-1)^{i+j} M_{ik}(A).$$

Proof. To be filled! See [CPY96, p.110 Proposition IV.23] ☐

Definition 6.3.3 (Cofactor). *We call cofactor of the coefficient a_{ij} in A, the scalar defined by:*

$$C_{ij}(A) = (-1)^{i+j} M_{ij}(A).$$

Remark 6.3.4 (Rules concerning signs). *The coefficient $(-1)^{i+j}$ equals 1 or -1.*

Theorem 6.3.5. *Let $A := (a_{ij})_{1\leqslant i,j\leqslant p}$ be a matrix of $\mathscr{M}_p(\mathbf{K})$. We have the following equalities:*

$$\boldsymbol{d}et(A) = \sum_{j=1}^{p} a_{ij} C_{ij}(A) \text{ (expanding along the } i^{th} \text{ row)},$$

$$\boldsymbol{d}et(A) = \sum_{i=1}^{p} a_{ij} C_{ij}(A) \text{ (expanding along the } j^{th} \text{ column)}.$$

Proof. To be filled! See [CPY96, p.112 Theoreme IV.25] ☐

Corollary 6.3.6. *If* $\det(A) \neq 0$, *then*

$$A^{-1} = \frac{1}{det(A)}\,{}^{t}(\tilde{A})$$

where $\tilde{A}$ *is the matrix of cofactors i.e.* $\tilde{A}_{ij} = C_{ij}(A)$, *for every* (i,j) *in* $[\![1,p]\!]$.

Proof. To be filled! See [CPY96, p.112 Corollaire IV.26] ☐

6.4 Computing a Determinant in practice

- ***Determinant of a triangular matrix***

  If $A := (a_{ij})_{1\leqslant i,j\leqslant p}$ *is a lower triangular matrix of* $\mathscr{M}_p(\mathbf{K})$, *by expanding, successively along the firs row, one gets:*

$$\begin{vmatrix} a_{1,1} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ a_{p,1} & \cdots & a_{p-1,p-1} & 0 \\ a_{p,1} & \cdots & a_{p,p-1} & a_{p,p} \end{vmatrix} = a_{1,1} \begin{vmatrix} a_{2,2} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ a_{p-1,2} & \cdots & a_{p-1,p-1} & 0 \\ a_{p,2} & \cdots & a_{p,p-1} & a_{p,p} \end{vmatrix} = a_{1,1}\cdots a_{p,p}.$$

  In view of Theorem 6.2.4, it is clear that the determinant of an upper triangular matrix equals the product of its diagonal terms.

- ***Make zeros appear***

  When the matrix is not triangular, "one can make appear zeros" by adding to a row (or a column) a linear combination of the others. It will not modify the value of the determinant. One can also move these zeros by exchanging two rows or two columns; of course this will modify the sign of the determinant. One can also multiply a row or a column by a scalar, which multiplies the determinant by this very scalar.

  All these operations remind the elementary operations of the Gaussian Elimination. The difference here is that, for the computation of a determinant one can work on rows as well as on columns as the same time.

  By using some of the elementary transformations described above one can reduce the computation of a determinant to the case where the matrix is triangular. This will therefore achieves the computation in view of the first point above.

- ***Expansion***

  The general method (i.e. the one described above) gives good results for numerical matrices (i.e. matrices with numbers instead of letters). However one can often improve it using the expansion along a row or a column when not many non zero coefficients left in this very row or column.

  When one wants to compute a determinant, denoted let's say Δ_p *which has a particular form (with* p *rows and* p *columns and* p *being unspecified), one can use this way of computing determinant in order to get a recursive formula which links* Δ_p *and* Δ_{p-1}.

- ***Numerical Techniques***

  In order to compute a determinant using a computer, we use methods adapted to each specific case. The LU factorization is a good example of such a method.

Definition 6.4.1 (). *Let $A := (a_{ij})$ be a matrix of $\mathscr{M}_{p,q}(\mathbf{K})$. We call extracted determinant (of A) of order $r \leqslant \min\{p, q\}$ the determinant of a square matrix of order r "extracted" from A. In other words by canceling $p - r$ rows and $q - r$ columns in the matrix A.*

Corollary 6.4.2. *The rank of a matrix is the greatest order of an extracted determinant different from 0.*

Proof. To be filled! See [CPY96, p.114 Corollaire IV.30] □

Theorem 6.4.3. *The rank of a matrix equals the rank of its transpose.*

Proof. To be filled! See [CPY96, p.115 Corollary IV.31] □

Part IV
Diagonalization

Chapter 7

Diagonalization

7.1 Eigenvalues of an Endomorphism

Definitions

Definition 7.1.1 (Eigenvalue and Eigenvector). *Let u be an endomorphism of a $\mathbf{K}$ vector space E. We say that a scalar λ is an eigenvalue of u if there exists a* ***non zero vector $\boldsymbol{x}$*** *of E such that:*

$$u(\boldsymbol{x}) = \lambda\, \boldsymbol{x}.$$

Such a vector $\boldsymbol{x}$ is called an eigenvector of u associated to the eigenvalue λ.

Remark 7.1.2. *Since, whatever the scalar λ, we have $u(0) = 0 = \lambda\, 0$, the requirement $\boldsymbol{x} \neq 0$ is essential in the previous definition.*

Definition 7.1.3 (Eigenvalue Space). *Let u be an endomorphism of a $\mathbf{K}$ vector space E and let λ be an eigenvalue of u. The vector subspace $E_\lambda := \ker(u - \lambda I_E)$, i.e. the set of all vectors $\boldsymbol{x}$ of E such that $u(\boldsymbol{x}) = \lambda\, \boldsymbol{x}$, is called the eigenspace of u associated with the eigenvalue λ.*

Example 7.1.4. *Let $\mathscr{B} := (\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$ be the standard basis of $\mathbf{R}^3$ and let $f : \mathbf{R}^3 \to \mathbf{R}^3$ be a linear map defined by setting:*

$$\begin{cases} f(\boldsymbol{e}_1) := \frac{-1}{3}\boldsymbol{e}_1 + \frac{2}{3}\boldsymbol{e}_2 + \frac{2}{3}\boldsymbol{e}_3 \\ f(\boldsymbol{e}_2) := \frac{2}{3}\boldsymbol{e}_1 + \frac{-1}{3}\boldsymbol{e}_2 + \frac{2}{3}\boldsymbol{e}_3 \\ f(\boldsymbol{e}_3) := \frac{2}{3}\boldsymbol{e}_1 + \frac{2}{3}\boldsymbol{e}_2 + \frac{-1}{3}\boldsymbol{e}_3. \end{cases}$$

Define $E_{-1} := \{\boldsymbol{u} \in \mathbf{R}^3,\ f(\boldsymbol{u}) = -\boldsymbol{u}\}$ and $E_1 := \{\boldsymbol{u} \in \mathbf{R}^3,\ f(\boldsymbol{u}) = \boldsymbol{u}\}$.

1. *Prove that $\boldsymbol{e}_1 - \boldsymbol{e}_2$ and $\boldsymbol{e}_1 - \boldsymbol{e}_3$ both belong to E_{-1} while $\boldsymbol{e}_1 + \boldsymbol{e}_2 + \boldsymbol{e}_3$ belongs to E_1.*

2. *Deduce that $\boldsymbol{e}_1 - \boldsymbol{e}_2, \boldsymbol{e}_1 - \boldsymbol{e}_3$ and $\boldsymbol{e}_1 + \boldsymbol{e}_2 + \boldsymbol{e}_3$ are eigenvalues of u.*

1. *Define $\boldsymbol{u} := \boldsymbol{e}_1 - \boldsymbol{e}_2$, $\boldsymbol{v} := \boldsymbol{e}_1 - \boldsymbol{e}_3$ and $\boldsymbol{w} := \boldsymbol{e}_1 + \boldsymbol{e}_2 + \boldsymbol{e}_3$. One can obviously write:*

$$\begin{aligned} f(\boldsymbol{u}) = f(\boldsymbol{e}_1 - \boldsymbol{e}_2) = f(\boldsymbol{e}_1) - f(\boldsymbol{e}_2) = \frac{-1}{3}\boldsymbol{e}_1 + \frac{2}{3}\boldsymbol{e}_2 + \frac{2}{3}\boldsymbol{e}_3 - (\frac{2}{3}\boldsymbol{e}_1 + \frac{-1}{3}\boldsymbol{e}_2 + \frac{2}{3}\boldsymbol{e}_3) = -\boldsymbol{e}_1 + \boldsymbol{e}_2 \\ = -\boldsymbol{u}. \end{aligned}$$

and

$$f(\boldsymbol{v}) = f(\boldsymbol{e}_1 - \boldsymbol{e}_3) = f(\boldsymbol{e}_1) - f(\boldsymbol{e}_2) = \frac{-1}{3}\boldsymbol{e}_1 + \frac{2}{3}\boldsymbol{e}_2 + \frac{2}{3}\boldsymbol{e}_3 - (\frac{2}{3}\boldsymbol{e}_1 + \frac{2}{3}\boldsymbol{e}_2 + \frac{-1}{3}\boldsymbol{e}_3) = -\boldsymbol{e}_1 + \boldsymbol{e}_3$$
$$= -\boldsymbol{v}.$$

These two equalities prove that both $\boldsymbol{u}$ and $\boldsymbol{v}$ belong to E_{-1}.

Moreover, we have:

$$f(\boldsymbol{e}_1 + \boldsymbol{e}_2 + \boldsymbol{e}_3) = f(\boldsymbol{e}_1) + f(\boldsymbol{e}_2) + f(\boldsymbol{e}_3)$$
$$= \frac{-1}{3}\boldsymbol{e}_1 + \frac{2}{3}\boldsymbol{e}_2 + \frac{2}{3}\boldsymbol{e}_3 + \frac{2}{3}\boldsymbol{e}_1 + \frac{-1}{3}\boldsymbol{e}_2 + \frac{2}{3}\boldsymbol{e}_3 + \frac{2}{3}\boldsymbol{e}_1 + \frac{2}{3}\boldsymbol{e}_2 + \frac{-1}{3}\boldsymbol{e}_3 = \boldsymbol{e}_1 + \boldsymbol{e}_2 + \boldsymbol{e}_3.$$

Thus $\boldsymbol{w}$ belongs to E_1.

2. *From the previous question we deduce that both $\boldsymbol{u}$ and $\boldsymbol{v}$ are eigenvectors associated to the eigenvalue -1. Moreover E_{-1} is the eigenspace associated to the eigenvalue -1.*

   Besides, and that $\boldsymbol{w}$ is an eigenvector associated to the eigenvalue 1. and E_1 is the eigenspace associated to the eigenvalue 1.

Proposition 7.1.5. *Let E be a finite-dimensional $\mathbf{K}$-vector space with $\dim(E) = p$ and let u be an endomorphism of E. The scalar $\boldsymbol{det}([u]_{\boldsymbol{e}}^{\boldsymbol{e}})$ does not depend on the basis $\boldsymbol{e}$ of E. We denote it $\boldsymbol{det}(u)$ and is called the determinant of the endomorphism u.*

Proof. To be filled! See [CPY96, p.136 Proposition V.3] □

Proposition 7.1.6. *Let E be a $\mathbf{K}$-finite-dimensional vector space with $\dim(E) = p$, u an endomorphism of E and λ be an element of $\mathbf{K}$. The following conditions are equivalent.*

1. *λ is an eigenvalue of u.*
2. *the endomorphism $u - \lambda I_E$ is not injective.*
3. *$\boldsymbol{det}(u - \lambda I_E)$=0.*

Proof. To be filled! See [CPY96, p.136 Proposition V.4] □

Definition 7.1.7 (Characteristic Polynomial). *Let E be a $\mathbf{K}$-finite-dimensional vector space with $\dim(E) = p$ and u be an endomorphism of E. Let us choose a basis $\boldsymbol{e}$ of E. It is easy to see that the quantity:*

$$\chi_u(X) = \boldsymbol{det}(u - XI_E) = \boldsymbol{det}([u - XI_E]_{\boldsymbol{e}}^{\boldsymbol{e}}) = \boldsymbol{det}([u]_{\boldsymbol{e}}^{\boldsymbol{e}} - XI_E)$$

is a polynomial in X. Its degree equals $p = \dim(E)$. This polynomial is called the Characteristic polynomial.

Applications to the Calculation of Eigenvalues

One computes first the characteristic polynomial χ_u, then one choose a basis $\boldsymbol{e}$ of E and one computes the determinant of the matrix $[u]_{\boldsymbol{e}}^{\boldsymbol{e}} - XI_E$. The eigenvalues of u are the roots, in $\mathbf{K}$, of the polynomial χ_u.

Example 7.1.8. *Let us keep the notations of Example 7.1.4.*

$$\chi_u(X) = \boldsymbol{det}([u]_{\boldsymbol{e}}^{\boldsymbol{e}} - XI_E) = \boldsymbol{det}(A - XI_3),$$

where we set $A := \begin{pmatrix} \frac{-1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{-1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{-1}{3} \end{pmatrix}$. *We hence can write:*

$$\chi_u(X) = \begin{vmatrix} X+\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & X+\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & X+\frac{1}{3} \end{vmatrix} = \begin{vmatrix} X+\frac{5}{3} & \frac{2}{3} & \frac{2}{3} \\ X+\frac{5}{3} & X+\frac{1}{3} & \frac{2}{3} \\ X+\frac{5}{3} & \frac{2}{3} & X+\frac{1}{3} \end{vmatrix} = (X+5/3)\cdot \begin{vmatrix} 1 & \frac{2}{3} & \frac{2}{3} \\ 1 & X+\frac{1}{3} & \frac{2}{3} \\ 1 & \frac{2}{3} & X+\frac{1}{3} \end{vmatrix}$$

$$= (X+5/3)\cdot \begin{vmatrix} 1 & \frac{2}{3} & \frac{2}{3} \\ 0 & X-\frac{1}{3} & 0 \\ 0 & 0 & X-\frac{1}{3} \end{vmatrix} = (X+5/3)\cdot \begin{vmatrix} X-\frac{1}{3} & 0 \\ 0 & X-\frac{1}{3} \end{vmatrix} = (X+5/3)\cdot (X-1/3)^2.$$

7.2 Diagonalizable Endomorphisms

Definition 7.2.1 (Diagonalizable Endomorphism). *An endomorphism u, of a finite dimensional vector space E, is said to be diagonalizable if there exists a basis $\boldsymbol{f}$ of E such that the matrix $[u]_{\boldsymbol{f}}^{\boldsymbol{f}}$ for which the matrix is diagonal.*
Diagonalize an endomorphism u means finding a basis $\boldsymbol{f}$ of E for which the matrix $[u]_{\boldsymbol{f}}^{\boldsymbol{f}}$ is diagonal. Of course this is possible only if the endomorphism u is diagonalizable.

Proposition 7.2.2. *Let E be a $\mathbf{K}$-finite-dimensional vector space with $\dim(E) = p$, u be an endomorphism of E and let $\boldsymbol{f} := (\boldsymbol{f}_1, \cdots, \boldsymbol{f}_p)$ be a basis of E. The matrix $[u]_{\boldsymbol{f}}^{\boldsymbol{f}}$ is diagonal if, and only if, the vectors $\boldsymbol{f}_i$ are eigenvectors of u. The i^{th} coefficient of its diagonal is then the eigenvalue of u associated to the corresponding eigenvector $\boldsymbol{f}_i$.*

Proof. To be filled! See [CPY96, p.138 Proposition V.8] □

Proposition 7.2.3. *If an an endomorphism u of a $\mathbf{K}$-finite-dimensional vector space E, with $\dim(E) = p$ is diagonalizable, its characteristic polynomial χ_u splits over $\mathbf{K}$ with no multiple zeros in $\mathbf{K}$ (there are p roots of χ_u in $\mathbf{K}$. Each of these roots is counted which its multiplicity).*

Proof. To be filled! See [CPY96, p.139 Proposition V.9] □

Lemma 7.2.4. *eddedujLet u be an endomorphism of a $\mathbf{K}$-finite-dimensional vector space E, with $\dim(E) = p$ and $\lambda_1, \cdots, \lambda_m$ some distinct eigenvalues of u and $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m$ eigenvectors of u associated to $\lambda_1, \cdots, \lambda_m$ such that $u(\boldsymbol{x}_i) = \lambda_i \boldsymbol{x}_i$ (i.e. $\boldsymbol{x}_i \in \ker(u - \lambda_i I_E)$).*
If $\boldsymbol{x}_1 + \cdots + \boldsymbol{x}_m = 0$ then $\boldsymbol{x}_1 = \cdots = \boldsymbol{x}_m = 0$.

Proof. To be filled! See [CPY96, p.139 Proposition V.10] □

Remark 7.2.5. *The previous result means that, for distinct eigenvalues $\lambda_1, \cdots, \lambda_m$, the corresponding eigenspaces $E_{\lambda_i} := \ker(u - \lambda_i I_E)$ all are in direct sum.*

Theorem 7.2.6. *Let u be an endomorphism of a $\mathbf{K}$-finite-dimensional vector space E, with $\dim(E) = p$. If the characteristic polynomial of u has p distinct roots in $\mathbf{K}$, i.e. if u has p eigenvalues, then u is diagonalizable.*

Proof. To be filled! See [CPY96, p.140 Theorème V.12] □

Proposition 7.2.7. *Let u be an endomorphism of a $\mathbf{K}$-finite-dimensional vector space E and let λ be an eigenvalue of u. Denote r the multiplicity order of λ as a root of χ_u. We have the following inequality:*

$$\dim(\ker(u - \lambda I_E)) \leqslant r.$$

Proof. To be filled! See [CPY96, p.140 Proposition V.13] □

Proposition 7.2.8. *Let u be an endomorphism of a $\mathbf{K}$-finite-dimensional vector space E with $\dim(E) = p$. Denote $\{\lambda_1, \cdots, \lambda_m\}$ the set of the eigenvalues (all distinct from each other) and, for every index i in $[\![1, m]\!]$, denote r_i the multiplicity of λ_i as a root of χ_u. The endomorphism u is diagonalizable if, and only if, one the following conditions hold:*

1. *The characteristic polynomial χ_u splits in $\mathbf{K}$ (i.e. it factors completely in $\mathbf{K}$) and for all i in $[\![1, m]\!]$, the equality $\dim(\ker(u - \lambda_i I_E)) = r_i$ holds.*
2. $\displaystyle\sum_{i=1}^{m} \dim(\ker(u - \lambda_i I_E)) = p.$
3. $E = \ker(u - \lambda_1 I_E) + \cdots + \ker(u - \lambda_m I_E).$

One can get a basis of E by gathering all the basis of every eigenspace $\ker(u - \lambda_i I_E)$.

Proof. To be filled! See [CPY96, p.140 Proposition V.14] □

7.3 Diagonalizable Matrices

Definition and Results

Definition 7.3.1 (Diagonalizable Matrix). *A square matrix A of $\mathscr{M}_p(\mathbf{K})$ is said to be diagonalizable (on $\mathbf{K}$) if the associated endomorphism $u : \mathbf{K}^p \to \mathbf{K}^p$, defined by settiing $X \mapsto AX$, is diagonalizable.*

We will keep, for matrices, the terminology we used for endomorphisms. We therefore call eigenvalue (resp. eigenvector) of A every eigenvalue (resp. every eigenvector) of the associated endomorphism u. The kernel of $u - \lambda I_E$, denoted $\ker(A - \lambda I_p)$, is called eigenspace of A, associated to the eigenvalue λ, and χ_A defined by setting $\chi_A(X) := \boldsymbol{det}(A - XI_p)$, is called the characteristic polynomial of A.

Theorem 7.3.2. *A matrix A of $\mathscr{M}_p(\mathbf{K})$ is diagonalizable if, and only if there exists in $\mathscr{M}_p(\mathbf{K})$, an invertible matrix P and a diagonal matrix D such that:*

$$D = P^{-1}AP. \tag{1}$$

Proof. To be filled! See [CPY96, p.141 Theorème V.16] □

Remark 7.3.3. *Matrices A and D are similar. Thus a matrix is diagonalizable if, and only if, it is similar to a diagonale matrix.*

Diagonalization in practice

Whenever A in $\mathscr{M}_p(\mathbf{K})$, is diagonalizable, diagonalize it means finding an invertible matrix P such that the matrix D, defined by setting:

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

is diagonal. Such a matrix P is the change of basis matrix from the standard basis $\boldsymbol{e}$ (which is the former basis) to a new basis, denoted $\boldsymbol{f}$, constituted of eigenvectors of A. The matrix D represents the endomorphism u in the basis $\boldsymbol{f}$. In other words, $D = [u]_{\boldsymbol{f}}^{\boldsymbol{f}}$.
Starting from matrix A, we get matrices D and P by the following manner.

1. *The eigenvalues of the matrix A are the roots of the characteristic polynomial χ_A (where $\chi_A(X) = \boldsymbol{d}et(A - XI_p)$).*

2. *We determine the eigenspaces of A and we choose a basis $\boldsymbol{f} := (\boldsymbol{f}_1, \cdots, \boldsymbol{f}_p)$ of $\mathbf{K}^p$ constituted of eigenvectors of A (we obtain it by taking the union of the basis of the eigenspaces). The i^{th} column of the matrix P is then constituted of the components of $\boldsymbol{f}_i$ in the standard basis of $\mathbf{K}^p$.*

3. *We have $D = diag(\lambda_1, \cdot, \lambda_p)$ where the λ_i are the eigenvectors of the matrix A. Each and every one of them appear on the diagonal of D, a number of times which equals its multiplicity, as a root of χ_A. They are ordered such that $\boldsymbol{f}_i$ is associated to λ_i.*

Example 7.3.4. *Denote $A := \begin{pmatrix} 3 & 4 \\ 5 & 2 \end{pmatrix}$ the matrix we want to diagonalize.*

1. *We start by looking for the roots of its characteristic polynomial.*

$$\chi_u(X) = \begin{vmatrix} 3-X & 4 \\ 5 & 2-X \end{vmatrix} = (3-X)(2-X) - 20 = X^2 - 5X - 14.$$

   It is easy to determine that the roots of χ_u are -2 and 7. Thus one can write:

$$\chi_u(X) = (X+2)(X-7).$$

   Since the characteristic polynomial of A has $p = 2$ distinct roots in $\mathbf{K}$, we know, according to Theorem 7.2.6 or Proposition 7.2.8, that A is diagonalizable.

2. *We are now looking for the eigenvectors of A. By definition, we know that the two eigenvectors are: $E_{-2} := \mathrm{Ker}(A + 2I_2)$ and $E_7 := \mathrm{Ker}(A - 7I_2)$.*

- *Make E_{-2} explicit*

  Let (x, y) be in $\mathbf{R}^2$.

$$^t(x,y) \in E_{-2} \iff \begin{pmatrix} 5 & 4 \\ 5 & 4 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \iff 5x + 4y = 0 \iff y = -\frac{5}{4}x$$

  One then concludes that $E_{-2} = \{(x, -\frac{5}{4}x),\ x \in \mathbf{R}\} = \mathrm{Span}_{\mathbf{R}}\{(-4,5)\}$. We can choose $\boldsymbol{x}_1 := (-4, 5)$ as an eigenvector of A, associated to the eigenvalue -2. Note moreover that $(-4, 5)$ is a basis of E_{-2}.

- *Make E_7 explicit*

  Let (x, y) be in $\mathbf{R}^2$.

$$^t(x,y) \in E_{7} \iff \begin{pmatrix} -4 & 4 \\ 5 & -5 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \iff -4x + 4y = 0 \iff y = x$$

  One then concludes that $E_7 = \{(x, x),\ x \in \mathbf{R}\} = \mathrm{Span}_{\mathbf{R}}\{(1,1)\}$. We can choose $\boldsymbol{x}_2 := {}^t(1,1)$ as an eigenvector of A, associated to the eigenvalue 7. Note moreover that $(1, 1)$ is a basis of E_7.

Finally gathering the two basis of the eigenspaces, we know that $\mathscr{B} := (\boldsymbol{x}_1, \boldsymbol{x}_2)$ is a basis of $\mathbf{R}^2$, constituted only with eigenvectors of A. Besides, define $D := [u]_{\mathscr{B}}^{\mathscr{B}}$. By definition of the $\boldsymbol{x}_i$, we know that $u(\boldsymbol{x}_1) = -2\boldsymbol{x}_1$ and that $u(\boldsymbol{x}_2) = 7\boldsymbol{x}_2$. It is therefore obvious that

$$D = [\boldsymbol{u}]_{\mathscr{B}}^{\mathscr{B}} = \begin{array}{cc} \begin{array}{cc} u(\boldsymbol{x}_1) & u(\boldsymbol{x}_2) \end{array} & \\ \begin{pmatrix} -2 & \quad 0 \\ 0 & \quad 7 \end{pmatrix} & \begin{array}{c} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{array} \end{array}.$$

Moreover, denote P the change of basis matrix from $\boldsymbol{e}$ (i.e. the standard basis of $\mathbf{R}^2$) and $\mathscr{B}$. By definition of P, one has the equality:

$$P = \begin{array}{cc} \begin{array}{cc} \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{array} & \\ \begin{pmatrix} -4 & 1 \\ 5 & 1 \end{pmatrix} & \begin{array}{c} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \end{array} \end{array}.$$

Since $A = [\boldsymbol{u}]_{\boldsymbol{e}}^{\boldsymbol{e}}$, we know, thanks to (9), that

$$D = P^{-1}AP. \tag{2}$$

One can here verify that Equality holds (2) by first computing P^{-1} and then by computing $P^{-1}AP$. We easily get

$$P^{-1} = \begin{pmatrix} -1/9 & 1/9 \\ 5/9 & 4/9 \end{pmatrix}$$

and then

$$P^{-1}AP = \begin{pmatrix} -1/9 & 1/9 \\ 5/9 & 4/9 \end{pmatrix} \cdot \begin{pmatrix} 3 & 4 \\ 5 & 2 \end{pmatrix} \cdot \begin{pmatrix} -4 & 1 \\ 5 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 0 \\ 0 & 7 \end{pmatrix}$$

which is nothing but D. We therefore recover the result which states that Equality (2) holds.

7.4 Applications

Computation of the n^{th} power of a matrix

Let A be diagonalizable matrix. Equality (1) also reads

$$A = PDP^{-1}. \tag{3}$$

It is therefore easy to deduce, that, for every n in $\mathbf{N}$,

$$A^n = PDP^{-1}PDP^{-1}\cdots PDP^{-1} = PD^nP^{-1}. \tag{4}$$

Matrix D (resp. A) represents the endomorphism u in the basis $\boldsymbol{f}$, which is constituted with eigenvectors of u (resp. in the standard basis $\mathbf{e}$). Equality (4) expresses simply the fact that both matrices A^n and D^n represent the same endomorphism (namely u^n), respectively in the basis $\mathbf{e}$ and in the basis $\boldsymbol{f}$.
Note moreover that the product of two diagonal matrices is very easy to compute. indeed, the product matrix is nothing but a diagonal matrix, the i^{th} term of which is the product of the i^{th} terms of both matrices. We therefore deduce that the diagonal matrix D^n is obtained by taking the n^{th} power of each and every of the factors on the diagonal of D.
Many other applications of diagonalization will be seen in Exercises. Note however that Diagonalization can also be successfully used to solve differential linear system with constant coefficients, to study system of Constant-recursive sequences.
J'en suis à [CPY96, p.143 Applications]

Part V

Multilinear Algebra

Chapter 8

Isometries

Chapter 9

Euclidean Spaces

Part VI

Applications

1. *PCA (ACP)*

2. *voir Strang et tout et tout!*

Part VII
Appendix

Maps, Image and Preimage

.1 Maps & Applications

Definition .1.1 (Map). *Let E and F be two non-empty sets. A map or a function f, from E to F is an assignment of exactly one element of F to each element of E. E is called Domain of f and F is called co-domain of f.*
If y_0 is the unique element of F assigned by the function f to the element x_0 of E, it is written as $f(x_0) = y_0$. f maps E to F means that f is a function from E to F, it is written as $f : E \to F$.

Example .1.2. *Let g, h, φ and ψ be defined by setting:*

$$g : \begin{array}{ccc} \mathbf{R} & \to & \mathbf{R} \\ x & \mapsto & x^2 \end{array} \qquad h : \begin{array}{ccc} \mathbf{R} & \to & \mathbf{R}_-^* \\ x & \mapsto & x^2 \end{array} \qquad \varphi : \begin{array}{ccc} \mathbf{R}_+ & \to & \mathbf{R}_+ \\ x & \mapsto & x^2 \end{array} \qquad \psi : \begin{array}{ccc} \mathbf{R} & \to & \mathbf{R}_+ \\ x & \mapsto & x^2 \end{array}$$

It is clear that the maps g, φ and ψ are correctly defined while h is not a map since it is impossible for the square of a real number to be negative.

Terms related to functions

Let

$$f : E \to F \tag{1}$$

be a map. Let us see the terms related to the definition of a function.
Domain and co-domain:
If f is a function from set E to set F, then E is called Domain and B is called co-domain.
Range:
The range of f is the set of all images of elements of E. Basically the range is subset of the co-domain. One can define it by

$$\textit{Range(f)} := \{f(u),\ u \in E\}.$$

Image and Pre-Image:
For any (a, b) in $E \times F$ (which means that a belongs to E and b belongs to F), b is the image of a and a is the pre-image of b if $f(a) = b$.
One can extend these last definitions to the cases of sets. More precisely, if we still consider the map f introduced in (1), *one can define:*

- *for every subset A of E the set*

$$f(A) := \{f(u),\ u \in A\}, \tag{2}$$

which is called the image of A by f. Note that $f(A)$ is a subset of F.

- *for every subset B of F the set*

$$f^{-1}(B) := \{u \in E,\ f(u) \in B\}, \tag{3}$$

which is called the preimage of B by f. Note that $f^{-1}(B)$ is a subset of E.

Remark .1.3. *From these previous definitions, one easily sees that the range of function is noting but the image of its domain.*

.2 Injectivity, surjectivity and bijectivity

There are three particular types of functions that are of great interest.

Definition .2.1 (One to one or injective function). *Let E and F be two non-empty sets. A function $f : E \to F$ is called one to one (or injective) if, for all elements x and x' in E, if $f(x) = f(x')$, then $x = x'$. It never maps distinct elements of its domain to the same element of its co-domain.*

Example .2.2. *If one considers the maps g, h, φ and ψ defined in Example .1.2, it is easy to see that:*

- *g is not injective since $g(-3) = 9 = g(3)$.*
- *φ is injective since, for all (a, b) in ${\mathbf{R}_+}^2$, one has:*

$$\varphi(a) = \varphi(b) \Longrightarrow a^2 = b^2 \Longrightarrow (a-b)\,(a+b) = 0 \Longrightarrow a = b \text{ or } a = -b.$$

  Since both a and b are positive, the case $a = -b$ is impossible and thus we conclude that

$$\forall (a, b) \in {\mathbf{R}_+}^2,\ \varphi(a) = \varphi(b) \Longrightarrow a = b.$$

  This proves the injectivity of φ.
- *ψ is not injective since $\psi(-3) = 9 = \psi(3)$.*

Definition .2.3 (Onto or surjective function). *Let E and F be two non-empty sets. A function $f : E \to F$ is called onto or surjective if all element of the codomain has a preimage in the domain. In other words f is surjective if, for all y of F there exists an element x in E such that $y = f(x)$.*
The function f may map one or more elements of E to the same element of F.

Example .2.4. *If one still considers the maps g, h, φ and ψ defined in Example .1.2, it is easy to see that:*

- *g is not surjective since, e.g.-4, does not have a preimage. Indeed, $-4 = x^2$ has no real solution. For the exact same reason h is not surjective either.*
- *Both φ and ψ are surjective since all b in $\mathbf{R}_+$ has a preimage in $\mathbf{R}_+$. Indeed, $\psi(\sqrt{b}) = b = \varphi(\sqrt{b})$, for all b in $\mathbf{R}_+$.*

We have the following result

Theorem .2.5. *Let $f : E \to F$ be a map. The two following statements are equivalent:*

1. *f is surjective*
2. *$f(E) = F$*

Proof. To be written $\square$

Definition .2.6 (One to one correspondence or Bijective function). *Let E and F be two non-empty sets. A function $f : E \to F$ is called bijective if it is both one to one and onto function (i.e. both injective and surjective function).*
For such a function, there exists a map, denoted $f^{-1} : F \to E$, characterized by the folloing equalities:

$$\forall x \in E,\ f^{-1}(f(x)) = x, \tag{4}$$

$$\forall y \in F,\ f(f^{-1}(y)) = y. \tag{5}$$

We usually call f^{-1} the inverse function of f.

Example .2.7. *If one still considers the maps g, h, φ and ψ defined in Example .1.2, it is clear, in view of Examples .2.2 and .2.4, that only φ is bijective. Its inverse is the map $\varphi^{-1} : \mathbf{R}_+ \to \mathbf{R}_+$ defined by: $\varphi^{-1}(u) := \sqrt{u}$, for every u in $\mathbf{R}_+$.*

Remark .2.8. *We have to be extremely cautious with the notation f^{-1}. Indeed, this notation can be used in two very different situations. Either one wants to speak about:*

- *the subset $f^{-1}(B)$ of E, where B is a subset of F (as defined in (3)). Such a set always exists (i.e. we do not need f to be bijective).*
- *the inverse function of f, which is possible only if f is bijective.*

Bibliography

[CPY96] *Gilles Christol, Philippe Pilibossian, and Sleiman Yammine.* Algèbre 2 - Espaces vectoriels, matrices, systèmes linéaires. *Ellipses, 1996.*

[HJ13] *Roger A. Horn and Charles R. Johnson.* Matrix Analysis. *Cambridge University Press, Cambridge, second edition, 2013.*

[OS18] *Peter J. Olver and Chehrzad Shakiban.* Applied Linear Algebra. *Undergraduate Texts in Mathematics. Springer, Cham, second edition, 2018.*