Linear algebra : a general overview

Guillaume Euvrard - May 2017

Warning : this document is not a class support. It has only been written in order to help my students at EPITA to review the main concepts of linear algebra and some of the typical situations they will meet. It should not be displayed out of the groups I work with.
In particular, this document has not been faced to a checking and validation process like a class support is expected to be. It hence may contain some remaining mistakes. Any remark from the readers that could help me to improve it will be wellcome.

1 Vector space over $\mathbb{R}$

The basic $\mathbb{R}$-vector spaces are
- $\mathbb{R}^n$,
- $\mathscr{M}_{n,p}(\mathbb{R})$, the set of matrices with n lines and p columns,
- $\mathbb{R}[X]$, the set of polynomials with real coefficients,
- $\mathbb{R}^{\mathbb{N}}$, the set of numerical sequences,
- $\mathbb{R}^I$, the set of the functions $I \to \mathbb{R}$ where I is an arbitrary given set.

Now, if E is a vector space over $\mathbb{R}$ and if $F \subset E$ is a subset of E, it is a vector space over $\mathbb{R}$ iff it is a linear subspace of E, i.e. :

$$\begin{cases} 0_E \in F \\ \forall (u,v) \in F^2, \forall \lambda \in \mathbb{R}, \lambda u + v \in F \end{cases}$$

It is important to have a precise idea about what 0_E is. It is not the real number 0 but the null vector of E, and its nature depends on E. For example, if $E = \mathbb{R}^I$, then 0_E is the null function over I :

$$\begin{array}{rccc} 0_E : & I & \to & \mathbb{R} \\ & x & \mapsto & 0 \end{array}$$

Thus, for any $f \in \mathbb{R}^I$, we have

$$f = 0_E \iff \forall x \in I, f(x) = 0$$

Particular case : finite dimensional vector space

Definitions

If E is a $\mathbb{R}$-vector space and if $B = (e_1, e_2, \ldots, e_n)$ is a family of E, we say
- B is *linearly independant* if

$$\forall (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n, \lambda_1 e_1 + \cdots + \lambda_n e_n = 0_E \Longrightarrow \lambda_1 = \cdots = \lambda_n = 0$$

It is *linearly dependant* if it is not linearly independant, i.e. :

$$\exists(\lambda_1,\dots,\lambda_n)\in\mathbb{R}^n, \lambda_1 e_1+\dots+\lambda_n e_n = 0_E \text{ and } (\lambda_1,\dots,\lambda_n)\neq(0,\dots,0)$$

— B is a *spanning family* of E if any element of E is a linear combination of B :

$$\forall u\in E, \exists(\lambda_1,\dots,\lambda_n)\in\mathbb{R}^n, u=\lambda_1 e_1+\dots+\lambda_n e_n$$

Remark : B is always a spanning family of $\mathrm{Span}(B)$, by definition of $\mathrm{Span}(B)$.

— B is a *basis* of E if it is both linearly independant and a spanning family of E.
Remark : if B is linearly independant, it is a basis of $\mathrm{Span}(B)$.
Remark 2 : if B is not linearly independant, we can find a subfamily B' of B which is linearly independant and such that $\mathrm{Span}(B')=\mathrm{Span}(B)$. We hence obtain a basis of $\mathrm{Span}(B)$.

Examples

— If $E=\mathbb{R}^n$, the family

$$\left(\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix},\begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix},\dots,\begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}\right)$$

is a basis of E. It is the standard basis of $\mathbb{R}^n$.

— The family $(1,X,X^2,\dots,X^n)$ is the standard basis of $\mathbb{R}_n[X]$.

— If $E=\mathbb{R}^{\mathbb{R}}$, the family $(f:x\mapsto e^x, g:x\mapsto e^{2x})$ is linearly independant. Indeed, for all $(\lambda_1,\lambda_2)\in\mathbb{R}^2$, we have

$$\begin{array}{rcl}\lambda_1 f+\lambda_2 g=0_E & \iff & \forall x\in\mathbb{R}, \lambda_1 f(x)+\lambda_2 g(x)=0\\ & \implies & \left\{\begin{array}{rcl}\lambda_1 f(0)+\lambda_2 g(0) & = & 0\\ \lambda_1 f(1)+\lambda_2 g(1) & = & 0\end{array}\right.\end{array}$$

Solving the latter system leads to the unique solution $(\lambda_1,\lambda_2)=(0,0)$ and the family (f,g) is hence linearly independant. It is a basis of $\mathrm{Span}(f,g)$.
Be careful, in the above deduction chain, the $\implies$ is not a $\iff$. If it happens that, for some functions f and g, the final system has another solution than $(\lambda_1,\lambda_2)=(0,0)$, it will not prove that (f,g) is linearly dependant. To reach this conclusion, you should find a relation $\lambda_1 f(x)+\lambda_2 g(x)=0$ which holds **for any** $x\in\mathbb{R}$.

— The family $(f:x\mapsto e^x, g:x\mapsto e^{x+1})$ is linearly dependant in $\mathbb{R}^{\mathbb{R}}$. Indeed, **for any** $x\in\mathbb{R}$, we have $ef(x)-g(x)=0$, which leads to

$$ef-g=0_E$$

Properties

We have the following proporties :

— If a vector space E has a basis, all other bases have the same number of vectors. This number is $\dim(E)$.

— If $\dim(E)=n$ and if B is a family of n vectors, then

$$B \text{ linearly independant} \iff B \text{ spanning family of } E \iff B \text{ basis of } E$$

So you only need to prove one of the two properties to prove that B is a basis.

— If F and G are two linear subspaces of E, then

$$F \subset G \Longrightarrow \dim(F) \leqslant \dim(G)$$

and

$$\left.\begin{array}{c} F \subset G \\ \dim(F) = \dim(G) \end{array}\right\} \Longrightarrow F = G$$

In particular $\dim(F) = \dim(E) \Longleftrightarrow F = E$, and $\dim(F) = 0 \Longleftrightarrow F = \{0_E\}$.

A consequence of these properties is that, if $\dim(E) = n$, then

— Any linearly independant family B of E has at most n vectors.
Proof : let's denote by p the number of vectors in B. As B is a basis of $\mathrm{Span}(B)$, we have

$$\left.\begin{array}{c} \dim\big(\mathrm{Span}(B)\big) = p \\ \mathrm{Span}(B) \subset E \end{array}\right\} \Longrightarrow p \leqslant \dim(E) = n$$

— Any spanning family B of E has at least n vectors.
Proof : let's denote by p the number of vectors in B. If B is linearly independant, then B is a basis of E and $p = n$. Else, we know we can select a subfamily B' of B which is linearly independant and such that $\mathrm{Span}(B') = \mathrm{Span}(B) = E$. This subfamily is hence a basis of E, so it is made of n vectors. Therefore, $n \leqslant p$.

2 Sum of subspaces

If E is a $\mathbb{R}$-vector space and if F and G are two linear subspaces of E, then $F+G$ is the set of all vectors of E that can be written as the sum of a vector of F and a vector of G :

$$F + G = \{w \in E, \exists (u,v) \in F \times G, w = u + v\}$$

If H is another subspace of E, the proof of $F + G = H$ requires the proof of both inclusions :

$\subset$: $\forall (u,v) \in F \times G, u + v \in H$

$\supset$: $\forall w \in H, \exists (u,v) \in F \times G, w = u + v$

If you need to prove $F + G = E$, the first inclusion is obvious but it's good to mention it. The second inclusion is usually the most difficult one to prove.

The sum $F+G$ is *direct* if $F \cap G = \{0_E\}$. Then we note $F \oplus G$. The proof of $F \cap G = \{0_E\}$ also requires both inclusions :

$\supset$: since F and G are linear subspaces, we have $0_E \in F$ and $0_E \in G$, so $\{0_E\} \subset F \cap G$,

$\subset$: let $u \in F \cap G$, then

$$\left.\begin{array}{c} u \in F \Longrightarrow \cdots \\ u \in G \Longrightarrow \cdots \end{array}\right\} \Longrightarrow \cdots \Longrightarrow u = 0_E$$

(we use the specific properties of F and G to do the deductions).

We say F and G are supplementary in E if $F \oplus G = E$.

Particular case of finite dimension

We have the relation

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

In particular, if the sum $F+G$ is direct, then $F\cap G=\{0_E\}$ has dimension 0 and we have

$$\dim(F\oplus G)=\dim(F)+\dim(G)$$

These relations may give some useful dimension arguments.
The most typical case is when we know that $\dim(F)+\dim(G)=\dim(E)$. Then, if we can prove $F\cap G=\{0_E\}$, we can deduce

$$\left.\begin{array}{rcl}\dim(F\oplus G) & = & \dim(F)+\dim(G)\\ & = & \dim(E)\\ F\oplus G & \subset & E\end{array}\right\}\Longrightarrow F\oplus G=E$$

3 Linear maps

If E and F are two vector spaces over $\mathbb{R}$, an application $f:E\mapsto F$ is a *linear map* from E to F, and we note it $f\in\mathscr{L}(E,F)$, if

$$\left\{\ \forall(u,v)\in E\times E,\forall\lambda\in\mathbb{R},f(\lambda u+v)=\lambda f(u)+f(v)\right.$$

A consequence of this definition is that for any $f\in\mathscr{L}(E,F)$, we have $f(0_E)=0_F$.
We say f is an *endomorphism* of E if its output space is E. We note it $f\in\mathscr{L}(E)$.

If $f\in\mathscr{L}(E,F)$, we define its kernel and its image :

— $\mathrm{Ker}(f)$ is the set of all the antecedants of 0_F :

$$\mathrm{Ker}(f)=\{u\in E,f(u)=0_F\}$$

It is a linear subspace of the *input space* E,

— $\mathrm{Im}(f)$ is the set of all the vectors of F which have at least one antecedant :

$$\mathrm{Im}(f)=\{v\in F,\exists u\in E,v=f(u)\}$$

It is a linear subspace of the *output space* F.

The kernel and image are related to the properties of injectivity and surjectivity :

— An application f is *injective* if any element of the output set has at most one antecedant by f. With the quantifiers, it is

$$\forall(x_1,x_2)\in E^2,f(x_1)=f(x_2)\Longrightarrow x_1=x_2$$

Now, if f is a linear map, we have

$$f\text{ injective}\Longleftrightarrow\mathrm{Ker}(f)=\{0_E\}$$

The inclusion $\supset$ is true for any linear map, injective or not, because $f(0_E)=0_F$. So the strong property for an injective map is the inclusion $\mathrm{Ker}(f)\subset\{0_E\}:\forall u\in E,f(u)=0_F\Longrightarrow u=0_E$.

— An application f is *surjective* if any element of F has at least one antecedant by f. This means $\mathrm{Im}(f)=F$.

Particular case of finite dimension

Matrix representation of a linear map

If E is a vector space of finite dimension n with a basis $B = (u_1, \cdots, u_n)$, and if F has dimension p with a basis $B' = (v_1, \cdots, v_p)$, then any linear map f from E to F may be represented with a matrix $\mathrm{Mat}_{BB'}(f)$. This matrix is defined as follows :

$$\begin{array}{c} \left(\begin{array}{ccc} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pn} \end{array}\right) \begin{array}{l} \leftarrow \text{coordinate for } v_1 \\ \\ \leftarrow \text{coordinate for } v_p \end{array} \\ \begin{array}{ccc} \uparrow & & \uparrow \\ f(u_1) & & f(u_n) \end{array} \end{array}$$

The i^{th} column of $\mathrm{Mat}_{BB'}(f)$ is made of the coordinates of $f(u_i)$ in the basis B'. Of course, this matrix depends on the bases B and B'.
Then, for any $x \in E$, if we denote by X and Y the coordinates columns of x and $f(x)$ in the bases B and B', we have

$$Y = \mathrm{Mat}_{BB'}(f) \times X$$

Precision : if $F = E$ and $B = B'$, then we note $\mathrm{Mat}_B(f)$ instead of $\mathrm{Mat}_{BB}(f)$.

The matrix product is consistant with the composition of linear maps :

$$\mathrm{Mat}_{BB''}(g \circ f) = \mathrm{Mat}_{B'B''}(g) \times \mathrm{Mat}_{BB'}(f)$$

Example : let $E = \mathbb{R}_2[X]$ and let $B = (1, X, X^2)$ and $B' = (1 - X, 1 + X, (1 + X)^2)$ be two bases of E. We consider the endormorphism f of E defined for all $P \in E$ by $f(P) = XP' - P$. Compute $\mathrm{Mat}_{BB'}(f)$.

Remark : the text of the exercise explicitly states that f is an endomorphism of E and that B' is a basis of E, so you don't need to prove it. Anyway, it is a good idea to check that you would know how to prove it if it was required.

Each column of $\mathrm{Mat}_{BB'}(f)$ is given by the image of the elements of B :

— the first column is given by $f(1)$: if $P = 1$, we have $P' = 0$ and $f(P) = XP' - P = -1$. Now, we need to write the polynomial -1 in the basis B' :

$$-1 = \alpha(1 - X) + \beta(1 + X) + \gamma(1 + X)^2$$

By expanding the right hand side of the latter expression and by identifying each coefficient, we find

$$f(1) = -1 = -\frac{1}{2}(1 - X) - \frac{1}{2}(1 + X) + 0(1 + X)^2$$

and the first column of the matrix is

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

— in the same way, for $P = X$, we find $f(P) = 0_E$ (the null polynomial) so the second column is made of 0s,
— and for $P = X^2, P' = 2X$ so $f(P) = X.2X - X^2 = X^2$. We now compute the coordinates of this polynomial in B' :

$$X^2 = \frac{1}{2}(1-X) - \frac{3}{2}(1+X) + (1+X)^2$$

which gives the third column.

Therefore,

$$\mathrm{Mat}_{BB'}(f) = \begin{pmatrix} -\frac{1}{2} & 0 & \frac{1}{2} \\ -\frac{1}{2} & 0 & -\frac{3}{2} \\ 0 & 0 & 1 \end{pmatrix}$$

The interpretation of this matrix is that, if we denote by (a_0, a_1, a_2) the coordinates in B of a polynomial P, (that is, $P = a_0 + a_1X + a_2X^2$), and if we denote by (b_0, b_1, b_2) the coordinates in B' of $f(P)$ (that is, $f(P) = b_0(1-X) + b_1(1+X) + b_2(1+X)^2$), then

$$\begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & 0 & \frac{1}{2} \\ -\frac{1}{2} & 0 & -\frac{3}{2} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}a_0 + \frac{1}{2}a_2 \\ -\frac{1}{2}a_0 - \frac{3}{2}a_2 \\ a_2 \end{pmatrix}$$

In other words,

$$f(a_0 + a_1X + a_2X^2) = \frac{-a_0 + a_2}{2}(1-X) - \frac{a_0 + 3a_2}{2}(1+X) + a_2(1+X)^2$$

Rank theorem

Let E and F be two finite dimensional vector spaces and $f \in \mathscr{L}(E,F)$. If $B = (u_1, \cdots, u_n)$ is a basis of E, then $f(B) = \big(f(u_1), \cdots, f(u_n)\big)$ is a spanning family of $\mathrm{Im}(f)$. It may happen that this family is not linearly independant, so we can conclude that $\dim\big(\mathrm{Im}(f)\big) \leqslant n = \dim(E)$. Rank theorem states that

$$\underbrace{\dim(E)}_{\text{input space}} = \dim\big(\mathrm{Ker}(f)\big) + \dim\big(\mathrm{Im}(f)\big)$$

This theorem is very helpful to give dimension arguments in a proof. For example, it enables to prove that :

— if $\dim(E) \neq \dim(F)$, the map f cannot be bijective :
when $\dim(E) < \dim(F)$, f cannot be surjective
when $\dim(E) > \dim(F)$, f cannot be injective
— if $\dim(E) = \dim(F)$, then

$$f \text{ injective} \iff f \text{ surjective} \iff f \text{ bijective}$$

Example 1 : let $E = \mathbb{R}_2[X]$ and f be defined for any $P \in E$ by $f(P) = XP' - P$. Give a basis of $\mathrm{Ker}(f)$ and a basis of $\mathrm{Im}(f)$. Check that rank theorem is satisfied.

Example 2 : let E be a finite dimensional vector space and $f \in \mathscr{L}(E)$ such that $f \circ f = f$ (we say that f is a projector). Show that $E = \mathrm{Ker}(f) \oplus \mathrm{Im}(f)$.
We need to prove two properties :

Prop 1 : $\mathrm{Ker}(f) \cap \mathrm{Im}(f) = \{0_E\}$.

$\supset$: since $\mathrm{Ker}(f)$ and $\mathrm{Im}(f)$ are linear subspaces of E, they both contain 0_E so $\{0_E\} \subset \mathrm{Ker}(f) \cap \mathrm{Im}(f)$

$\subset$: let $u \in \mathrm{Ker}(f) \cap \mathrm{Im}(f)$. Then

$$\left\{ \begin{array}{lcl} u \in \mathrm{Ker}(f) & \Longrightarrow & f(u) = 0_E \\ u \in \mathrm{Im}(f) & \Longrightarrow & \exists v \in E, u = f(v) \end{array} \right.$$

If we inject $u = f(v)$ in the equation $f(u) = 0_E$, we get

$$\begin{array}{rcll} f(u) = 0_E & \Longrightarrow & f \circ f(v) = 0_E & \\ & \Longrightarrow & f(v) = 0_E & (\text{since } f \circ f = f) \\ & \Longrightarrow & u = 0_E & (\text{since } u = f(v)) \end{array}$$

so $\mathrm{Ker}(f) \cap \mathrm{Im}(f) \subset \{0_E\}$.

Prop 2 : $\mathrm{Ker}(f) \oplus \mathrm{Im}(f) = E$

$\subset$: since both $\mathrm{Ker}(f)$ and $\mathrm{Im}(f)$ are subspaces of E, we have $\mathrm{Ker}(f) \oplus \mathrm{Im}(f) \subset E$

$\supset$: we combine the dimension relation of a sum of subspaces with rank theorem :

$$\left\{ \begin{array}{lcl} \dim\big(\mathrm{Ker}(f)\big) + \dim\big(\mathrm{Im}(f)\big) & = & \dim\big(\mathrm{Ker}(f) + \mathrm{Im}(f)\big) + \dim\big(\mathrm{Ker}(f) \cap \mathrm{Im}(f)\big) \\ \dim\big(\mathrm{Ker}(f)\big) + \dim\big(\mathrm{Im}(f)\big) & = & \dim(E) \end{array} \right.$$

Now, since $\mathrm{Ker}(f) \cap \mathrm{Im}(f) = \{0_E\}$ has dimension 0, we can deduce that

$$\left. \begin{array}{rcl} \dim\big(\mathrm{Ker}(f) \oplus \mathrm{Im}(f)\big) & = & \dim(E) \\ \mathrm{Ker}(f) \oplus \mathrm{Im}(f) & \subset & E \end{array} \right\} \Longrightarrow \mathrm{Ker}(f) \oplus \mathrm{Im}(f) = E$$

Remark : the latter inclusion is the only part of the exercise where we used a dimension argument. In fact, in this particular example, it can be proven with another argument : for all $u \in E$, we have

$$u = \underbrace{u - f(u)}_{\in \mathrm{Ker}(f)} + \underbrace{f(u)}_{\in \mathrm{Im}(f)}$$

Thus, the property $E = \mathrm{Ker}(f) \oplus \mathrm{Im}(f)$ still holds in infinite dimension. But it is easier to prove it in finite dimension.

4 Polynomial of endomorphism and of square matrix

If $f \in \mathscr{L}(E)$, we can define f^n for any $n \in \mathbb{N}$ by

$$\left\{ \begin{array}{l} f^0 = id \\ f^n = \underbrace{f \circ f \circ \cdots \circ f}_{n \text{ times}} \text{ if } n > 0 \end{array} \right.$$

Then, for all $(p, q) \in \mathbb{N}^2$, f^p and f^q commute with the composition :

$$f^p \circ f^q = f^q \circ f^p = f^{p+q}$$

This property and the linearity of f imply that all the polynomial expressions of f commute, and that the usual way to multiply polynomials still holds when applied to f : if P and Q are polynomials, then

$$P(f) \circ Q(f) = Q(f) \circ P(f) = PQ(f)$$

where PQ is the polynomial obtained by multiplying P and Q.
For example, since $(X-1)(X-2) = X^2 - 3X + 2$, we have

$$(f - id) \circ (f - 2id) = (f - 2id) \circ (f - id) = f^2 - 3f + 2id$$

You will have hints and reminders in the text of the exercise if you're asked to do some polynomial reasonings refering to the first semester.

Particular case of finite dimension

When E is a finite dimensional vector space with a basis $B = (e_1, \cdots, e_n)$, then $f \in \mathscr{L}(E)$ can be represented by a square matrix

$$A = \mathrm{Mat}_B(f) \in \mathscr{M}_n(\mathbb{R})$$

Then all the above holds, when the composition law $\circ$ is replaced with the matrix product. For example,

$$(A - I_n)(A - 2I_n) = (A - 2I_n)(A - I_n) = A^2 - 3A + 2I_n$$

Careful : something that *we can not deduce* is

$$(A - I_n)(A - 2I_n) = 0 \underbrace{\Longrightarrow}_{\text{wrong}} A - I_n = 0 \quad \text{or} \quad A - 2I_n = 0$$

Example : let A be the matrix

$$A = \begin{pmatrix} 3 & -5 & 1 \\ 1 & -3 & 1 \\ 3 & -15 & 5 \end{pmatrix}$$

We can check that

$$(A - I_3)(A - 2I_3) = \begin{pmatrix} 2 & -5 & 1 \\ 1 & -4 & 1 \\ 3 & -15 & 4 \end{pmatrix} \begin{pmatrix} 1 & -5 & 1 \\ 1 & -5 & 1 \\ 3 & -15 & 3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Yet, we have neither $A - I_3 = 0$, nor $A - 2I_3 = 0$. According to an exercise done in the TD, the relation $(A - I_3)(A - 2I_3) = 0$ implies that

$$\mathbb{R}^3 = \mathrm{Ker}(A - I_3) \oplus \mathrm{Ker}(A - 2I_3)$$

This relation can also be used to deduce A^n for any $n \in \mathbb{N}$. To do this, we do the Euclidean division of the polynomial X^n by $(X-1)(X-2)$: there exist Q and R in $\mathbb{R}[X]$ such that

$$\begin{cases} X^n = Q(X)(X-1)(X-2) + R(X) \\ \deg(R) < 2 \end{cases}$$

Since $\deg(R) < 2$, there exists $(a, b) \in \mathbb{R}^2$ such that $R(X) = aX + b$ and we just need to identify the coefficients a and b. We do this identification with the particular values $x = 1$ and $x = 2$: since

$Q(X)(X-1)(X-2)$ cancels for these values, we have

$$X^n = Q(X)(X-1)(X-2) + aX + b \implies \begin{cases} 1^n &= a \times 1 + b \\ 2^n &= a \times 2 + b \end{cases}$$

$$\implies \begin{cases} a &= 2^n - 1 \\ b &= 2 - 2^n \end{cases}$$

We hence have

$$X^n = Q(X)(X-1)(X-2) + (2^n - 1)X + (2 - 2^n)$$

Therefore,

$$\begin{aligned} A^n &= Q(A)\underbrace{(A - I_3)(A - 2I_3)}_{=0} + (2^n - 1)A + (2 - 2^n)I_3 \\ &= (2^n - 1)A + (2 - 2^n)I_3 \end{aligned}$$