Contents

2/7/2017

1 Fields

First we introduce the sets that are going to serve as the generalizations of the scalars.

Definition 1.1 We say that a set K with two laws $+$ and $\times$ is a *field* if the following conditions hold :

(1) The law $+$ has the following properties :
   a) It's commutative : $a + b = b + a$;
   b) It's associative : $a + (b + c) = (a + b) + c$;
   c) It has an identity element 0 (which means that $0 + a = a + 0 = a$ for every $a \in K$).
   d) For every $a \in K$, there exists $b \in K$ such that $a + b = b + a = 0$. (We write $b = -a$.)

(2) The law $\times$ is also commutative ($a \times b = b \times a$) and associative ($a \times (b \times c) = (a \times b) \times c$), it also has an identity element, which is called a (so we have $a \times 1 = 1 \times a = a$ for every $a \in K$), and moreover $\times$ is distributive with respect to $+$, which means that : for every $a, b, c \in K$,

$$a \times (b + c) = (a \times b) + (a \times c) \quad \text{and} \quad (b + c) \times a = (b \times a) + (c \times a).$$

(3) For every $a \in K$ such that $a \neq 0$, there exists $b \in K$ such that $a \times b = b \times a = 1$. (We say that b is the inverse of a and write $b = a^{-1}$.)

(4) $0 \neq 1$.

Remark 1.2
- If K satisfies conditions 1 and 2 only, we say that K is a *commutative ring*. An example is the ring of polynomials $\mathbb{Q}[X]$. We could do linear algebra over commutative rings, but it's more complicated and is not the goal of this class.
- If K satisfies conditions 1 and 2, *except for the commutativity of* $\times$, we say that K is a *ring*. An example is the set square $n \times n$ matrices $M_n(\mathbb{Q})$. Rings are very interesting but are also not the focus of this class; we will only meet a few examples, so it's convenient to have a name.

- We often don't write the sign $\times$ (so we write ab instead of $a \times b$). We also usually write $a - b$ instead of $a + (-b)$.
- There are a few facts that we take for granted in usual arithmetic, such as the fact that $a \times 0 = 0$. In a general field, they are still true but we have to prove them before we can use them. To prove some of these facts effectively, the notion of a *group* will be useful, though it's not necessary to know what it is to understand linear algebra.

Example 1.3 Here are some examples (and non-examples).

- $\mathbb{N} := \mathbb{Z}_{\geq 0}$ is not a field or even a ring, because 2 (for example) has no additive inverse in $\mathbb{N}$.
- $\mathbb{Z}$ is not a field, becaues 2 has no multiplicative inverse in $\mathbb{Z}$. However, it is a commutative ring.
- $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are all field.
- Let $n \in \mathbb{Z}_{\geq 1}$. The set of integers modulo n, $\mathbb{Z}/n\mathbb{Z}$, [1] is a commutative ring. It is a field if and only if n is a prime number (see problem set 1). If $n = p$ is a primer number, we also write $\mathbb{F}_p$ for $\mathbb{Z}/p\mathbb{Z}$.

Definition 1.4 We say that a set G with one law $*$ is a *group* if the following conditions hold :

(1) $*$ is associative : $a * (b * c) = (a * b) * c$.

(2) $*$ admits an identity, denoted by e (so we have $a * e = e * a = a$ for every $a \in G$).

(3) For every $a \in G$, there exists $b \in G$ such that $a * b = b * a = e$. We say that b is the inverse of a.

If moreover $*$ is commutative ($a * b = b * a$), we say that G is a *commutative group* (or *abelian group*).

Lemma 1.5 *Let G be a group (with the law $*$). Then its identity element is unique, and, for every $a \in G$, the element b of (3) is also unique.*

Proof. Suppose that we have two identity elements, e and e'. Then, using the property of (2), we get $e = e * e' = e'$.

Let $a \in G$, and suppose that we have two elements $b, b' \in G$ such that $a * b = b * a = e$ and $a * b' = b' * a = e$. Then :

$$b = b * e = b * (a * b') = (b * a) * b' = e * b' = e.$$

□

[1] http://en.wikipedia.org/wiki/Modular_arithmetic

Example 1.6

- If K is a ring (in particular a field), the set K with the law $+$ is a commutative group. Its identity element is 0, and the inverse of a (for the addition) is $-a$.
- Let K be a ring (commutative or not). Let $K^\times$ be the set of a in K that admit an inverse (that is, such that there exist b in K with $ab = ba = 1$). The $K^\times$ with the law $\times$ is a group. The identity element is 1, and the inverse of a is a^{-1}.

  Note that, if K is a field, then $K^\times = K - \{0\}$.
- If X is a set, the set $\mathfrak{S}_X$ of bijections $u : X \to X$, together with the law of composition, is a group (it's not commutative if $|X| \geq 2$). If $X = \{1, \ldots, n\}$, we write $\mathfrak{S}_X = \mathfrak{S}_n$, and call it the symmetric group. This is just an example and you can forget it for now.

Lemma 1.7 *Let K be a field (or even just a ring). Then :*

(i) The identity element 0 of addition, the additive inverse $-a$ of $a \in K$, the multiplicative identity 1, and the multiplicative inverse a^{-1} of $a \in K$ (if it exists) are all uniquely determined.

(ii) For every $a, b \in K$, $a \times 0 = 0 \times a = 0$ and $a \times (-b) = -ab$.

Proof. Point (i) follows from the properties of groups that we proved above, and the examples following it. Let's prove property (ii). Fix $a, b \in K$. Then :

$$(a \times 0) = (a \times 0) + a - a = a \times 0 + (a \times 1) - a = a \times (0 + 1) - a = (a \times 1) - a = a - a = 0,$$

and

$$ab + a(-b) = a(b - b) = a \times 0 = 0.$$

□

2 Vector spaces

Definition 2.1 Let K be a field. A *vector space over K* (or *K-vector space*) is a set V together with two laws, an addition that takes two elements v, w of V and returns an element $v + w$ of V, and a scalar multiplication that takes an element a *of* K and an element v of V and returns an element $a \times v = av$ of V, satisfying the following properties :

(1) The set V with the operation $+$ is a commutative group (which means that $+$ is commutative and associative, that it has an identity element which we'll call 0, and that every $v \in V$ has an inverse for V, which we'll call $-v$).

(2) We have $1 \times v = v$ for every $v \in V$.

(3) For every $a, b \in K$ and $v \in V$, we have $a(bv) = (ab)v$.

(4) The law $\times$ is distributive with respect to the addition on K and on V, that is, for every $a, b \in K$ and $v, w \in V$:

$$(a+b)v = av + bv$$

$$a(v+w) = av + aw.$$

Note that we cannot multiply two elements of V ! We can only multiply an element of V by an element of K. Note also that are using the same notation 0 for the additive identities in K and in V, although those are two different objects.

Note also that property (1) is just saying that V with the law $+$ is a commutative group.

Very often, we call the elements of V *vectors* and the elements of K *scalars*.

Just as before, there are some properties that we are tempted to take for granted, but we actually have to prove them.

Lemma 2.2 *Let V be a K-vector space. Then the following hold :*

(i) For every $v \in V$, $0 \times v = 0$ and $(-1) \times v = -v$.

(ii) For every $a \in K$ and $v \in V$, $(-a)v = a(-v) = -(av)$.

(iii) For every $a \in K$, $a \times 0 = 0$.

In the future, we'll just write $-av$ for $-(av)$.

Proof. Let $a \in K$ and $v \in V$. We have :

$$0 \times v = 0 \times v + v - v = 0 \times v + 1 \times v - v = (0+1)v - v = 1 \times v - v = v - v = 0.$$

Then we get :

$$v + (-1) \times v = 1 \times v + (-1) \times v = (1-1) \times v = 0 \times v = 0.$$

This proves (i). To prove (ii), note that :

$$av + (-a)v = (a-a)v = 0 \times v = 0$$

(by (i)), so $(-a)v = -(av)$. On the other hand :

$$a(-v) = a((-1) \times v) = (a \times (-1))v = (-a)v,$$

and we have just proved that this is $-(av)$. Finally, for (iii), note that

$$a \times 0 = a \times (v - v) = av + a \times (-v) = av - av = 0.$$

□

Example 2.3 There are two basic examples of K-vector spaces. The first one is a particular case of the second one (take $I = \{1, \ldots, n\}$).

- If $n \in \mathbb{Z}_{\geq 1}$, the set K^n of ordered n-uples o$(x_1, \ldots, x_n)$ of elements of K is a K-vector space with the following two laws :
  - $(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$;
  - $a(x_1, \ldots, x_n) = (ax_1, \ldots, ax_n)$, if $a \in K$.

  For example, you might have met $\mathbb{R}^2$ and $\mathbb{R}^3$ before.

- Let I be a non-empty set. The set K^I of functions $f : I \to K$ is a K-vector space, with the following two laws :
  - If $f, g \in K^I$, $f + g \in K^I$ is defined by $(f + g)(i) = f(i) + g(i)$ for every $i \in I$.
  - If $f \in K^I$ and $a \in K$, $af \in K^I$ is defined by $(af)(i) = af(i)$ for every $i \in I$.

  Note that the additive identity in K^I is the constant function equal to 0. We denote it by 0, in agreement with our general convention.

Checking that these sets indeed satisfy all the properties is straightforward but wquite tedious. (You should do it as en exercise.) Usually, it's easier to check that something is a *subspace*, as in the following definition :

Definition 2.4 Let V be a K-vector space. We say that a subset W of V is a *K-subspace* (or just a subspace if K is clear) if we have :

(1) $0 \in W$;

(2) For every $v, w \in W$, $v + w \in W$.

(3) For every $v \in W$ and $a \in K$, $av \in W$.

Note that, by (3), if $v \in W$, then $-v = (-1)v$ is also in W.
The following is a straightforward-but-tedious verification and left as an exercise.

Lemma 2.5 *If V is a K-vector space and W is a K-subspace of V, then W is also a K-vector space.*

Example 2.6 If V is a K-vector space, then V itself and $\{0\}$ are subspaces of V.

2/9/2017

Example 2.7

- Take $K = \mathbb{R}$ and $V = \mathbb{R}^3$. Then the subset defined by $x_1 + x_2 + x_3 = 1$ is not a subspace (because it doesn't contain 0), but the subset defined by $2x_1 - x_2 + 3x_3 = 0$ is a subspace.

- In $\mathbb{R}$, $\mathbb{Q}$ is a $\mathbb{Q}$-subspace but not a $\mathbb{R}$-subspace.

- Let K be a field and I be a non-empty set, and consider the K-vector space K^I of last time. If $f \in K^I$, its *support* $\mathrm{supp}(f)$ is by definition the set of $i \in I$ such that $f(i) \neq 0$. Let $K^{(I)}$ be the subset of K^I of functions with finite support. Then $K^{(I)}$ is a K-subspace of K^I.

  Why is it a subspace ? This follows from the follwing three properties of the support (exercise) :

  - $\mathrm{supp}(0) = \varnothing$;
  - for every $f, g \in K^I$, $\mathrm{supp}(f+g) \subset \mathrm{supp}(f) \cup \mathrm{supp}(g)$;
  - for every $a \in K$ and $f \in K^I$, $\mathrm{supp}(af) \subset \mathrm{supp}(f)$.

- Here is a subexample of the previous example. Let's take $I = \mathbb{N} := \mathbb{Z}_{\geq 0}$. Then $K^{(\mathbb{N})}$ is sometimes written $K[X]$ and called the space of polynomials with coefficients in K (in one indeterminate). When we do this, the convention is that X^n denote tha function f that sends n to 1 and every $m \neq n$ to 0; we also write $1 = X^0$. Then a function $f \in K^{(\mathbb{N})}$ can be rewritten as $\sum_{n \in \mathbb{N}} f(n) X^n$ (the sum is finite because we assumed that the support of f is finite).

  We can also define a multiplication on $K[X]$, by setting $X^n X^m = X^{n+m}$ and extending this by distributivity. Then $K[X]$ becomes a commutative ring.

- Here is an example of vector space over the field $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$. Let I be a non-empty set, and take V to be the set of subsets of I, with the following operations :

  - If $A, B \in V$, then $A + B = (A \cup B) - (A \cap B)$.
  - If $A \in V$, then $0 \times A = \varnothing$ and $1 \times A = A$.

  This is a somewhat silly example, because it is actually the same as the vector space $\mathbb{F}_2^I$. We just have to identify the subset A with its characteristic function, that is, the function f_A defined by

$$f_A(i) = \begin{cases} 1 & \text{if } i \in A \\ 0 & \text{if } i \notin A. \end{cases}$$

  It's easy to check that the operations are the same.

3 Matrices

Let K be a field.

Definition 3.1 A $n \times m$ matrix with coefficients in K is a table of elements of K with n rows and m columns. We call these elements of K the *entries* of the matrix.

The set of $n \times m$ matrices with coeffiecients in K is denoted by $M_{nm}(k)$. If $n = m$, we write $M_n(K)$ instead of $M_{nn}(K)$ (and we talk about square matrices of size n).

There are two standard ways to refer to the entries of a matrix. Either we write "let $A = (x_{ij})$ be a $n \times m$ matrix", then this means that the entry in position (i, j) is called

x_{ij}. Or we just write "let A be a $n \times m$ matrix", then the entry in position (i, j) is usually called A_{ij}.

Now let's define some operations on matrices.

Addition and scalar multiplication

If A and B are in $M_{nm}(K)$ and $x \in K$, we define :

- the sum $A + B \in M_{nm}(K)$ by $(A + B)_{ij} = A_{ij} + B_{ij}$;
- the product $xA \in M_{nm}(K)$ by $(xA)_{ij} = xA_{ij}$.

This makes $M_{nm}(K)$ into a K-vector space, in fact, it's the same vector space as K^{nm}. (For example, as vector spaces, $M_{23}(K) = K^6$.)

In particular, both $M_{n1}(K)$ and $M_{1n}(K)$ can be identified with K^n. We call the elements of $M_{n1}(K)$ *column vectors* and write them $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. We call the elements of $M_{1n}(K)$ *row vectors* and write them $\begin{pmatrix} x_1 & \dots & x_n \end{pmatrix}$. Usually we think of K^n as the space of column vectors.

Transpose

Definition 3.2 If $A \in M_{nm}(K)$, its *transpose*, denoted by ${}^tA$ or A^T, is the $m \times n$ matrix given by $(A^T)_{ij} = A_{ji}$.

Lemma 3.3 *Let $A, B \in M_{nm}(K)$ and $a \in K$. Then :*

(i) $(A^T)^T = A$.

(ii) $(A + B)^T = A^T + B^T$.

(iii) $(aA)^T = aA^T$.

Proof. Points (ii) and (iii) are obvious verifications. For (ii), note that $(A^T)^T$ is a $n \times m$ matrix (the same size as A), and that we have, for every $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$:

$$((A^T)^T)_{ij} = (A^T)_{ji} = A_{ij}.$$

So $(A^T)^T = A$.

□

Note that, if $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ is a column vector, then $v^T = \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix}$ is a row vector. (And vice versa.)

Matrix multiplication

Definition 3.4 Let $A \in M_{nm}(K)$ and $B \in M_{mp}(K)$. (Note the sizes : the number of columns of A is equal to the number of rows of B.) Then their product AB is the $n \times p$ matrix defined by :

$$(AB)_{ij} = \sum_{r=1}^{m} A_{ir} B_{rj},$$

for every $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$.

We have the following properties of matrix multiplication (all very easy and proved in class) :

Lemma 3.5 *Suppose that $A, B \in M_{nm}(K)$, $C, D \in M_{mp}(K)$, and $E \in M_{pq}(K)$.*

(i) $(A + B)C = AC + BC$.

(ii) $A(C + D) = AC + AD$.

(iii) $(AC)E = A(CE)$.

Definition 3.6 the *identity matrix* of size n is the square matrix I_n of size n defined by :

$$(I_n)_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

We also write $I_n = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$.

The following was also proved in class :

Lemma 3.7 *If $A \in M_{nm}(K)$, then $I_n A = A I_m = A$.*

Remark 3.8 In particular, in the space of square matrices $M_n(k)$, we have an addition and a multiplication that satisfy the following properties : addition is commutative and associative, it has an identity element and additive inverses exist; multiplication is associative, has an identity element and is distributive with respect to addition. This is what we called a ring (although you don't need to remember this).

Remark 3.9 Here are some properties that multiplication of matrices does *not* have :

(1) It is *not* commutative. First, if $A \in M_{nm}(K)$ and $B \in M_{mp}(K)$, then AB makes sense but BA does not make sense in general. Suppose that $p = n$, then BA makes sense, but it is square of size m while AB is square of size n, so if $n \neq m$ it does not make sense to ask if AB and BA are the same. Finally, assume that $p = n = m$, so that A and B are both square of size n. Then it is still not true in general that AB and BA are the same !

Here is an example : $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in M_2(\mathbb{Q})$, $B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \in M_2(\mathbb{Q})$. Then :

$$AB = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in M_2(\mathbb{Q}) \neq BA = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \in M_2(\mathbb{Q}).$$

(2) It is *not* true that general nonzero square n matrices have inverses. (And the question doesn't even make sense for non-square matrices.) For example, take $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in M_2(\mathbb{Q})$ and $B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \in M_2(\mathbb{Q})$ as in the previous example. We have seen that $BA = 0$. Suppose that A had an inverse, that means that there is a 2×2 matrix C such that $AC = CA = I_2$. Then we would have :

$$B = B(AC) = (BA)C = 0C = 0,$$

which is not true. So A doesn't have an inverse, even though it is nonzero.

4 Linear systems and matrices

Definition 4.1 Let $\vec{v}_1, \ldots, \vec{v}_m, \vec{w}$ be elements of K^n.

(1) We say that $\vec{w}$ is *a linear combination of* $\vec{v}_1, \ldots, \vec{v}_m$ if there exists scalars $\lambda_1, \ldots, \lambda_m \in K$ such that $\vec{w} = \lambda_1\vec{v}_1 + \cdots + \lambda_m\vec{v}_m$.

(2) We say that the vectors $\vec{v}_1, \ldots, \vec{v}_m$ are *linearly independent* (or *free*) if the only scalars $\lambda_1, \ldots, \lambda_m \in K$ such that $\lambda_1\vec{v}_1 + \cdots + \lambda_m\vec{v}_m = 0$ are $\lambda_1 = \cdots = \lambda_m = 0$. (In other words, $\lambda_1\vec{v}_1 + \cdots + \lambda_m\vec{v}_m = 0 \Leftrightarrow \lambda_1 = \cdots = \lambda_m = 0$.)

2/14/2017

Definition 4.2 (sequel)

(3) The *span* of the family $(\vec{v}_1, \ldots, \vec{v}_m)$ is the set of linear combinations of $\vec{v}_1, \ldots, \vec{v}_m$ (as in (1)). We denote this set by $\mathrm{Span}(\vec{v}_1, \ldots, \vec{v}_m)$. (It's easy to see that it's a subspace of K^n, and we will also give a proof of this later.)

Here is the connection with linear systems : Let $A = (a_{ij})$ be the $n \times m$ matrix whose columns are $\vec{v}_1, \ldots, \vec{v}_m$ (seen as column vectors), in that order. This means that $\vec{v}_j$ is the column vector $\begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix}$. We also see $\vec{w}$ as a column vector (= a $n \times 1$ matrix), and we write λ for the column vector $\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix}$ (this one is a $m \times 1$ matrix).

Then it follows directly from the definition that the matrix product $A\lambda$ is the $n \times 1$ matrix (= column vector) $\lambda_1\vec{v}_1 + \cdots + \lambda_m\vec{v}_m$: indeed, its entry at the position $(i, 1)$ is equal to

$$\sum_{j=1}^{m} a_{ij}\lambda_j = \sum_{j=1}^{m} \lambda_j a_{ij},$$

which is exactly the jth coordinate of $\lambda_1\vec{v}_1 + \cdots + \lambda_m\vec{v}_m$.

Now let $\vec{x} = \begin{pmatrix} v_1 \\ \vdots \\ x_m \end{pmatrix}$ be a column vector of unknowns. If $\vec{w} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$, then the matrix equation

$$A\vec{x} = \vec{w}$$

is just another (and more compact) way to write the following system of linear equations in the unknowns $x_1, \ldots, x_m$:

$$(S) \quad \begin{cases} a_{11}x_1 + \cdots + a_{1m}x_m &= b_1 \\ & \cdots \\ a_{n1}x_1 + \cdots + a_{nm}x_m &= b_n \end{cases}$$

We can reformulate the definitions above in the following way :

(1) The vector $\vec{w}$ is a linear combination of the vectors $\vec{v}_1, \ldots, \vec{v}_m$ if and only if the system (S) above has a least one solution.

(2) The vectors $\vec{v}_1, \ldots, \vec{v}_m$ are linearly independent if and only if the only solution of the system :

$$\begin{cases} a_{11}x_1 + \cdots + a_{1m}x_m &= 0 \\ & \cdots \\ a_{n1}x_1 + \cdots + a_{nm}x_m &= 0 \end{cases}$$

is $x_1 = x_2 = \cdots = x_m = 0$.

(3) The span of $\vec{v}_1, \ldots, \vec{v}_m$ is the set of $\vec{w} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$ such that the system (S) above has at least one solution.

So it is useful to know how to solve systems of linear equations.

Here are two other definitions (basically two other names for things we already know) :

Definition 4.3 Let A be a $n \times m$ matrix. As before, we denote by $\vec{v}_1, \ldots, \vec{v}_m$ the columns of A, seen as elements of K^n.

(1) The *image* of A, denoted by $\mathrm{Im}(A)$, is the subset of K^n made up of $\vec{w}$ such that the equation $A\vec{x} = \vec{w}$ has at lest one solution. In other words, it's just another name for the span of $\vec{v}_1, \ldots, \vec{v}_n$ (= the span of the columns of A).

(2) The *kernel* of A, denoted by $\mathrm{Ker}(A)$, is the subset of K^m made up of the vectors $\vec{\lambda} = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix}$ such that $A\vec{\lambda} = 0$. In other words, it's the set of solutions of the equation $A\vec{x} = 0$.

It's an easy exercise to show that $\mathrm{Im}(A)$ is a K-subspace of K^n and $\mathrm{Ker}(A)$ is a K-subspace of K^m. (This will come up again later.)

5 Resolution of linear systems

You have probably learned how to solve systems of linear equations before. The standard algorithm is to apply elementary operations on the rows, without forgetting the second term, until the system is in reduced two echelon form. Remember that "elementary row operations" are the following three operations : switching two rows, multiplying a row by a nonzero scalar, adding a multiple of a row to another row. We'll see shortly how to reinterpret them as matrix operations, but for now let's review reduced row echelon form.

Definition 5.1 We say that a matrix A is in *reduced row echelon form* if :

(1) Every row of A is of the form $\begin{pmatrix} 0 & \dots & 0 & 1 & * & \dots & * \end{pmatrix}$. That is, it starts with some number of 0's, then there's a 1, then we allow any scalars as entries. Note that we allow the row to start with the 1 directly (that is, the number of 0's in front can be zero), and we also allow the row to be all 0's.

(2) For every i, the first 1 on the row $i+1$ has to appear at a later position than the first 1 on the row i.

The 1's that start the rows of A are called the *pivots* of A, and the number of pivots is called the *rank* of A.[2]

Example 5.2 The following matrices are in reduced row echelon form :

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

The following matrices are not :

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

[2]This is not the standard definition of rank and will be superseded later.

Suppose that we want to solve a system of linear equations, write in matrix form $A\vec{x} = \vec{w}$ as above. The *augmented matrix* of the system is by definition the matrix $B = (A|\vec{w})$. (That is, the matrix A with the column vector $\vec{w}$ added as a last column.) To solve the system, we perform *elementary row operations* to put the augmented matrix B in reduced rwo echelon form. The point is that this does not change the set of solutions of the system, and that a system in reduced row echelon form is very easy to solve.

Definition 5.3 Here is a list of the three elementary row operations, and their matricial interpretation :

(1) Switching two rows of B : if we switch the rows number r and s of B, this is the same as multiplying B on the left by the $n \times n$ matrix C given by

$$C_{ij} = \begin{cases} 1 & \text{if } i = j \text{ and } i \neq r \text{ and } j \neq s \\ 1 & \text{if } (i,j) = (r,s) \text{ or } (i,j) = (s,r) \\ 0 & \text{otherwise} \end{cases}$$

Note that this matrix C is invertible, and in fact we have $C^2 = I_n$ (so C is its own inverse).

(2) Multiplying a row of B by a nonzero scalar : multiplying row number r of B by the nonzero scalar $a \in K$ is the same as multiplying the matrix B on the left by the $n \times n$ matrix C given by

$$C_{ij} = \begin{cases} 1 & \text{if } i = j \text{ and } i \neq r \\ a & \text{if } i = j = r \\ 0 & \text{if } i \neq j \end{cases}$$

Note that this matrix C is invertible, and in fact its inverse is the matrix C' defined like C but replacing a by a^{-1}.

(3) Adding a multiple of a row to another row : adding a times row number r to row number s of B (with $a \in K$ and $r \neq s$) is the same as multiplying B on the left by the matrix C given by

$$C_{ij} = \begin{cases} 1 & \text{if } i = j \\ a & \text{if } (i,j) = (s,r) \\ 0 & \text{otherwise} \end{cases}$$

Note this matrix C is invertible, and in fact its inverse is the matrix C' defined like C but replacing a by $-a$.

Theorem 5.4 *Let B be a $n \times p$ matrix. Then B can be put in reduced row echelon form after a finite number of elementary row operations. In other words, there exists a square matrix C of size n, product of matrices as in the description above, such that CB is in reduced row echelon form.*

We define the *rank* of B to be the rank of CB.[3]

Suppose that $B = (A|\vec{w})$ is the augmented matrix of the system $A\vec{x} = \vec{w}$ as above, and choose C as in the theorem. By the remarks above, C is invertible. So

$$A\vec{x} = \vec{w} \Leftrightarrow (CB)\vec{x} = C\vec{w},$$

which means that the systems with augmented matrices B and CB have the same solutions. This is why our method of solving systems works. Note that the system has solutions if and only the matrix CB has no rows of the form $\begin{pmatrix}0 & \dots & 0 & 1\end{pmatrix}$. If we are trying to find equations of the span of the columns of A, then the second term of the system, $\vec{w}$, is a vector of indeterminates, and the condition that we cannot put B in reduced row echelon form with a row of the form $\begin{pmatrix}0 & \dots & 0 & 1\end{pmatrix}$ will translate as some linear conditions on the entries of $\vec{w}$.

Now let's prove the theorem.

Proof. We prove the theorem by induction on the number n of rows of B.

If $n = 1$, then we multiply the only row of B by the inverse of the first nonzero coefficient of that row to put B in reduced row echelon form. This is an elementary row operation, so we're done.

Suppose that $n \geq 2$ and that we know the theorem for $n-1$. First, let $i_0 \in \{1, \dots, n\}$ be the number such that row number i_0 starts with the smallest number of 0's among all the rows of B. If $i_0 \neq 1$, we switch rows 1 and i_0, which is allowed. Then we multiply row 1 (the ex-row i_0) by the inverse of its first nonzero coeffiecient to make its first nonzero coefficient equal to 1. Suppose that the new matrix is equal to (d_{ij}), with $d_{11} = d_{12} = \cdots = d_{1,s-1} = 0$ and $d_{1s} = 1$ (s can be equal to 1). Then, for each $i \in \{1, \dots, n\}$, we replace row i by row i plus $(-d_{is})$ times row 1. This has the effect of killing all the cofficients in the sth column except for the one in the first row. Also, by the choice of i_0 above, the coefficients in columns $1, 2, \dots, s-1$ are all 0. This means that the new matrix B is of the form :

$$B = \begin{pmatrix} 0 & \dots & 0 & 1 & b \\ 0 & \ddots & 0 & 0 & B' \end{pmatrix},$$

where b is a row vector of size $1 \times (p-s-1)$ and B' is a matrix of size $(n-1) \times (p-s-1)$. To finish the proof, we apply the induction hypothesis to B'. (Nothing that elementary row operations on B' can be seen as elementary row operations on B that only affects rows 2 to n.)

□

Example 5.5 Suppose that we want to put the matrix $B = \begin{pmatrix} 0 & 0 & 3 & 1 & 1 \\ 0 & 2 & 0 & 7 & -1 \\ 0 & 1 & 0 & -3 & 0 \end{pmatrix}$ in reduced row echelon form. (This one is slightly different from the example done in

[3] Again, this is not the standard definition and will be superseded. It is also not clear a priori that the rank is well-defined. We will prove this later, when we give the more standard definition.

class.) Here is the sequence of operations described in the proof above (if we unpack the induction) :

- Switch rows 1 and 3 to get $\begin{pmatrix} 0 & 1 & 0 & -3 & 0 \\ 0 & 2 & 0 & 7 & -1 \\ 0 & 0 & 3 & 1 & 1 \end{pmatrix}$ (Note : there is a choice here, we could also have started by switching rows 1 and 2.)
- Add -2 times row 1 to row 2 (to make the first nonzero coefficient of row 2 be on the right of the leading 1 of row 1, as it should be). We get $\begin{pmatrix} 0 & 1 & 0 & -3 & 0 \\ 0 & 0 & 0 & 13 & -1 \\ 0 & 0 & 3 & 1 & 1 \end{pmatrix}$.
- Switch rows 2 and 3 to get $\begin{pmatrix} 0 & 1 & 0 & -3 & 0 \\ 0 & 0 & 3 & 1 & 1 \\ 0 & 0 & 0 & 13 & -1 \end{pmatrix}$.
- Multiply row 2 by $1/3$ to get $\begin{pmatrix} 0 & 1 & 0 & -3 & 0 \\ 0 & 0 & 1 & 1/3 & 1/3 \\ 0 & 0 & 0 & 13 & -1 \end{pmatrix}$.
- Multiply row 3 by $1/13$ to get $\begin{pmatrix} 0 & 1 & 0 & -3 & 0 \\ 0 & 0 & 1 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1 & -1/13 \end{pmatrix}$.

And we're done !

6 Calculating the inverse of a matrix using elementary row and column operations

Suppose that A is a square $n \times n$ matrix and that we want to decide if it's invertible, and to calculate its inverse if it is. There is a method to do this that is very similar to what we did to solve systems.

Consider the matrix $B = (A|I_n)$. (That's a $n \times (2n)$ matrix.) Applying the algorithm of the previous section gives an invertible matrix C such that $CB = (CA|C)$ is in reduced row echelon form. (Note that we do not have to keep track of C, since it will naturally appear as the right half of CB.)

If the last line of CA is $\begin{pmatrix} 0 & \dots & 0 \end{pmatrix}$, then the matrix A is not invertible. (Why ? Because, iff A were invertible, then CA would be invertible too. But a matrix whose last line is $\begin{pmatrix} 0 & \dots 0 \end{pmatrix}$ cannot be invertible. This is obvious if $n = 1$, because then the condition says that $CA = 0$. If $n = 2$, then multiplying CA on the right by the nonzero matrix $\begin{pmatrix} I_{n-1} & 0 \\ 0 & 0 \end{pmatrix}$ would give 0, and we have seen before that this prevents a matrix from being invertible.)

If the last row of CA is not $\begin{pmatrix} 0 & \dots & 0 \end{pmatrix}$, then CA is of the form $\begin{pmatrix} 1 & & * \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$.

Then I claim that we can perform some more elementary row operations on CB to pur it in the form $(I_n|*)$. That is, we can find another invertible matrix C' such that $C'(CB) = (I_n|C'C)$. (Note again that we do not need to keep track of $C'C$, the second half of the matrix does it for us.) And then $C'C$ is the inverse of A. Indeed, $C'(CB) = (C'C)B = ((C'C)A|C'C) = (I_n|C'C)$, so $(C'C)A = I_n$.[4]

So we just have to prove the following theorem to check that everything works :

Theorem 6.1 *If D is a $n \times n$ matrix of the form $\begin{pmatrix} 1 & & * \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$, then we can transform D into the identity matrix I_n by a finite number of elementary row operations.*

Proof. We reason by induction on n. If $n = 1$, then $D = I_1$ and there is nothing to prove.

Suppose that $n \geq 2$ and that we know the result for $n-1$. For every $i \in \{1, \dots, n-1\}$, we add $(-D_{in})$ times the last row to row numer i. This has the effect of making the last column of D equal to $\begin{pmatrix} 0 \\ \vdots \\ 1 \end{pmatrix}$, and so D becomes of the form $\begin{pmatrix} D' & 0 \\ 0 & 1 \end{pmatrix}$, where D' is a square matrix of size $n-1$ such that $D' = \begin{pmatrix} 1 & & * \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$. Then we conclude by applying the induction hypothesis to D'.

□

7 Linear combination and linear independence in general vector spaces

We still suppose that K is a field, and we fix a K-vector space V.

Definition 7.1 Let $\vec{v}_1, \dots, \vec{v}_n$ be elements of V.

(1) We say that $\vec{w} \in V$ is *a linear combination of* $\vec{v}_1, \dots, \vec{v}_m$ if there exists scalars $\lambda_1, \dots, \lambda_m \in K$ such that $\vec{w} = \lambda_1\vec{v}_1 + \cdots + \lambda_m\vec{v}_m$.

(2) We say that the vectors $\vec{v}_1, \dots, \vec{v}_m$ are *linearly independent* (or *free*) if the only scalars $\lambda_1, \dots, \lambda_m \in K$ such that $\lambda_1\vec{v}_1 + \cdots + \lambda_m\vec{v}_m = 0$ are $\lambda_1 = \cdots = \lambda_m = 0$. (In other words, $\lambda_1\vec{v}_1 + \cdots + \lambda_m\vec{v}_m = 0 \Leftrightarrow \lambda_1 = \cdots = \lambda_m = 0$.)

(3) The *span* of the family $(\vec{v}_1, \dots, \vec{v}_m)$ is the set of linear combinations of $\vec{v}_1, \dots, \vec{v}_m$ (as in (1)).

[4]Note that we have only show that $(C'C)A = I_n$, and normally we should also show that $A(C'C) = I_n$. Later, we will see that the second condition follows automatically from the first.

You might notice that this is exactly the same definition as in K^n (definition 4.1).

Example 7.2 Suppose that $V = K[X]$, and consider the family of vectors $(1, X, X^3)$. Then $2 - 3X + 4X^3$ is a linear combination of these vectors, but $1 - 2X + X^2$ is not. Also, the family $(1, X, X^3)$ is linearly independent (because a polynomial is 0 if and only if all its coefficients are 0.)

We also have a definition for a possibly infinite family of vectors. These are the only reasonable definitions, given that we can only form finite sums.

Definition 7.3 Let $(\vec{v}_i)_{i\in I}$ be a family of elements of V.

(1) We say that $\vec{w} \in V$ is *a linear combination the family* $(\vec{v}_i)_{i\in I}$ if there exists $m \geq 1$, $i_1, \ldots, i_m \in I$ and scalars $\lambda_1, \ldots, \lambda_m \in K$ such that $\vec{w} = \lambda_1\vec{v}_{i-1} + \cdots + \lambda_m\vec{v}_{i_m}$. (That is, if $\vec{w}$ is a linear combination of some finite subfamily of $(\vec{v}_{i\in I})$.

(2) We say that the family $(\vec{v}_i)_{i\in I}$ is *linearly independent* (or *free*) if every finite subfamily of $(\vec{v}_i)_{i\in I}$ is free.

(3) The *span* $\mathrm{Span}(\vec{v}_i, i \in I)$ of the family $(\vec{v}_i)_{i\in I}$ is the set of linear combinations of $(\vec{v}_i)_{i\in I}$ (as in (1)). We also say that $(\vec{v}_i)_{i\in I}$ is a generating family of $\mathrm{Span}(\vec{v}_i, i \in I)$.

By convention, the span of the empty family is $\{0\}$. (This is coherent with the standard convention that an empty sum should be $\{0\}$.)

2/16/2017

Example 7.4 In $V = K[X]$, consider the infinite family $(X^n)_{n\geq 0}$. (Remember that $X^0 = 1$.) Then this family is free (because a polynomial is 0 if and only if all its coefficients are 0), and its span is V itself (because every polynomial is by definition a linear combination of the X^n.)

It's also useful to think about the case of two or three vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$, what it means for them to be linearly independent and what the span is.

Proposition 7.5 *Let $(\vec{v}_i)_{i\in I}$ be a family of vectors in a K-vector space V.*

(i) The span of the family $(\vec{v}_i)_{i\in I}$ is the smallest K-subspace of V containing all the $\vec{v}_i$.

(ii) If $\vec{v} \in \mathrm{Span}(\vec{v}_i, i \in I)$, then $\mathrm{Span}(\vec{v}_i, i \in I) = \mathrm{Span}(\vec{v}, \vec{v}_i, i \in I)$.

(iii) If the family $(\vec{v}_i)_{i\in I}$ is free, and if $\vec{v} \notin \mathrm{Span}(\vec{v}_i, i \in I)$, then the family $(\vec{v}, \vec{v}_i, i \in I)$ is also free.

Proof.

(i) Let $W = \mathrm{Span}(\vec{v}_i, i \in I)$. First we must show that W is a K-subspace of V. There are three conditions to check :

- $0 \in W$: This is true because 0 is equal to the empty sum. [5]
- If $\vec{v}, \vec{w} \in W$, then $\vec{v} + \vec{w} \in W$: Indeed we can write $\vec{v} = \sum_{r=1}^{n} \lambda_r \vec{v}_{i_r}$ and $\vec{w} = \sum_{s=1}^{m} \mu_s \vec{v}_{j_s}$, with $\lambda_1, \ldots, \lambda_n, \mu_1, \ldots, \mu_m \in K$ and $i_1, \ldots, i_n, j_1, \ldots, j_m \in I$. Then $\vec{v} + \vec{w} = \sum_{r=1}^{n} \lambda_r \vec{v}_{i_r} + \sum_{s=1}^{m} \mu_s \vec{v}_{j_s}$ is also a linear combination of the $\vec{v}_i$, $i \in I$.
- If $\vec{v} \in W$ and $\lambda \in K$, then $\lambda\vec{v} \in W$: Indeed we can write $\vec{v} = \sum_{r=1}^{n} \lambda_r \vec{v}_{i_r}$ as before, and then $\lambda\vec{v} = \sum_{r=1}^{n} (\lambda\lambda_r)\vec{v}_{i_r}$, which is clearly still a linear combination of the $\vec{v}_i$, $i \in I$.

Now we also have to see that, if W' is another K-subspace of V such that $\vec{v}_i$ is in W' for every $i \in I$, then $W \subset W'$. Let $\vec{v}$ be an element of W. By definition of the span, we have $\vec{v} = \sum_{r=1}^{n} \lambda_r \vec{v}_{i_r}$ with $\lambda_1, \ldots, \lambda_n \in K$ and $i_1, \ldots, i_n \in I$. As $\vec{v}_{i_1}, \ldots, \vec{v}_{i_n} \in W'$ (by hypothesis), we have $\lambda_1\vec{v}_{i_1}, \ldots, \lambda_r\vec{v}_{i_n} \in W'$, and so their sum $\vec{v}$ is also in W'. As this is true for any element of W, we have shown that $W \subset W'$.

(ii) This follows from (i). (If $\vec{v} \in \mathrm{Span}(\vec{v}_i, i \in I)$, then $\mathrm{Span}(\vec{v}_i, i \in I)$ is a subspace containing $\vec{v}$ and all the $\vec{v}_i$, so it contains $\mathrm{Span}(\vec{v}, \vec{v}_i, i \in I)$.)

(iii) We prove the result by contradiction. Suppose that the family $(\vec{v}, \vec{v}_i, i \in I)$ is not free, then there exists $i_1, \ldots, i_n \in I$ and $\lambda, \lambda_1, \ldots, \lambda_n \in K$ such that $\lambda\vec{v} + \lambda_1\vec{v}_{i_1} + \cdots + \lambda_n\vec{i_n} = 0$ and that at least one of $\lambda, \lambda_1, \ldots, \lambda_n$ is nonzero. If $\lambda = 0$, then we get that $\lambda_1\vec{v}_{i_1} + \cdots + \lambda_n\vec{i_n} = 0$ and that at least of the λ_i is nonzero, which contradicts the freeness of the family $(\vec{v}_i)_{i\in I}$. So $\lambda \neq 0$. But then we have

$$\vec{v} = -\lambda^{-1}(\lambda_1\vec{v}_{i_1} + \cdots + \lambda_n\vec{i_n}),$$

which shows that $\vec{v} \in \mathrm{Span}(\vec{v}_i, i \in I)$, contradiction.

□

8 Bases

Definition 8.1 Let V be a K-vector space. A family $(\vec{v}_i)_{i\in I}$ of vectors of V is called *a basis of* V if it is free and its span is V.

Lemma 8.2 *Let V be a K-vector space. A family $\mathfrak{B} = (v_i)_{i\in I}$ of vectors of V is basis if and only if, for every $\vec{v} \in V$, there exists a* unique *family $(\lambda_i)_{i\in I}$ of elements of K such that :*

(a) All but a finite number of the λ_i are zero.

[5] This might sound like cheating. Another way to think about it is to say that, if I is not empty, then we can choose some $i \in I$ and then $0 = 0\vec{v}_i$. And remember that the span of the empty family is $\{0\}$ by convention.

(b) $\vec{v} = \sum_{i\in I} \lambda_i \vec{v}_i$ *(this sum is finite thanks to (a)).*

(In other words, if and only if every element of V *can be written as a linear combination of the* $\vec{v}_i$ in a unique way.*)*

The λ_i *are called the* coordinates of $\vec{v}$ in the basis $\mathfrak{B}$*, and we sometimes write* $(\lambda_i)_{i\in I} = [\vec{v}]_{\mathfrak{B}}$ *(this is usually seen as a column vector if* I *is finite).*

Proof. Suppose that $(\vec{v}_i)_{i\in I}$ is a basis of V, and let $\vec{v}$ be in V. As $(\vec{v}_i)_{i\in I}$, there exists a family $(\lambda_i)_{i\in I}$ satisfying (a) and (b). Suppose that we have another family $(\lambda'_i)_{i\in I}$ also satisfying (a) and (b), and let $\mu_i = \lambda_i - \lambda'_i$. Then all but a finite number of the μ_i are zero, so the sum $\sum_{i\in I} \mu_i \vec{v}_i$ makes sense, and this sum is equal to $\vec{v} - \vec{v} = 0$. As the family $(\vec{v}_{i\in I}$ is free, all the μ_i must be zero, which means that $\lambda_i = \lambda'_i$ for every $i \in I$.

Conversely, suppose that, for every $\vec{v} \in V$, we have a unique family $(\lambda_i)_{i\in I}$ satisfying (a) and (b). Then in particular every $\vec{v} \in V$ is a linear combination of the $\vec{v}_i$, so the family $(\vec{v}_i)_{i\in I}$ spans V. Let's show that it is free. Suppose that we have a relation $a_1\vec{v}_{i_1} + \cdots + a_n\vec{v}_{i_n} = 0$ (with $i_1, \ldots, i_n$ pairwise distinct). Then setting $\lambda_i = a_r$ for $i = i_r$ and $\lambda_i = 0$ for $i \notin \{i_1, \ldots, i_n\}$, we get a family $(\lambda_i)_{i\in I}$ satisfying (a) and (b) for $\vec{v} = 0$. By the uniqueness, this implies that $\lambda_i = 0$ for every $i \in I$, and in particular $a_1 = \cdots = a_n = 0$.

□

Example 8.3

- In K^n, the family of the vectors $\vec{e}_1 = (1, 0, \ldots, 0)$, $\vec{e}_2 = (0, 1, 0, \ldots, 0)$,$\ldots$,$\vec{e}_n = (0, \ldots, 0, 1)$ is a basis, called the *canonical basis* of K^n.

  To prove this, we just have need to notice that $\sum_{i=1}^n x_i e_i = (x_1, \ldots, x_n)$ if $x_1, \ldots, x_n \in K$, so every element of K^n is a linear combination of $\vec{e}_1, \ldots, \vec{e}_n$ in a unique way. (And in fact the coordinates of an element of K^n are its coordinates in the canonical basis.)

- In $K[X]$, the family $(1, X, X^2, X^3, \ldots)$ is a basis (also often called the canonical basis). The coordinates of a polynomial in this basis are its coefficients.

- More generally, let I be a set and consider the K-vector space $K^{(I)}$. For every $i \in I$, define an element $e_i \in K^{(I)}$ by setting

$$e_i(j) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases}$$

  Then $(e_i)_{i\in I}$ is a basis of $K^{(I)}$, still called the canonical basis. If $f \in K^{(I)}$, we have $f = \sum_{i\in I} f(i)e_i$, so the coordinates of f in the basis $(e_i)_{i\in I}$ are given by the family $(f(i))_{i\in I}$. [6]

- The family $(1, 2), (0, 1))$ is a basis of $\mathbb{R}^2$.

[6] Note that if we tried to use K^I instead of $K^{(I)}$, this would fail because we could get an infinite sum if we wrote $\sum_{i\in I} f(i)e_i$.

9 Dimension

Definition 9.1 We say that a K-vector V is *finite-dimensional* if it has a finite generating family (that is, if it can be be spanned by a finite family of vectors). Otherwise we say that V is *infinite-dimensional*.

If V is finite-dimensional, its *dimension* $\dim(V)$ is by definition the minimum of the cardinalities of the generating families of V.

We will see shortly that a finite-dimensional vector space always has a finite basis, and that the dimension is just the cardinality of a basis (and that all bases have the same cardinality).

Remark 9.2

The only 0-dimensional K-vector space is $\{0\}$.

Theorem 9.3 *Let V be a finite-dimensional K-vector space, and let $n = \dim(V)$.*

(i) If $(\vec{v}_i)_{i\in I}$ is a finite generating family of V, then there exists a subset J of I such that $(\vec{v}_i)_{i\in J}$ is a basis of V. (In other words, we can extract a basis from any finite generating family.) In particular, V admits finite bases.

(ii) Any basis of V has cardinality n.

Proof.

(i) If the family $(\vec{v}_i)_{i\in I}$ is free, then we are done.

Otherwise, there is a linear relation $\lambda_1\vec{v}_{i_1} + \cdots + \lambda_r\vec{v}_{i_r} = 0$, with $\lambda_1, \ldots, \lambda_r \in K$ not all zero, and $i_1, \ldots, i_r \in I$ pairwise distinct. Up to changing the numbering, we may assume that $\lambda_1 \neq 0$. Then

$$\vec{v}_{i_1} = -\lambda_1^{-1}(\lambda_2\vec{v}_{i_2} + \cdots + \lambda_r\vec{v}_{i_r},$$

so $\vec{v}_{i_1} \in \mathrm{Span}(\vec{v}_i, i \in I - \{i_1\})$, so $\mathrm{Span}(\vec{v}_i, i \in I - \{i_1\}) = \mathrm{Span}(\vec{v}_i, i \in I) = V$.

We repeat the previous step with the family $(\vec{v}_i)_{I-\{i_1\}}$, which has cardinality stricly smaller than $|I|$. Because I is finite, this process has to stop after a finite number of steps, and produces a subfamily of $(\vec{v}_i)_{i\in I}$ which is both free and generating, also known as a basis.

(ii) We reason by induction on n. The statement if empty if $n = 0$, so let's assume $n \geq 1$. Fix a basis $(\vec{e}_1, \ldots, \vec{e}_n)$ of V (such a basis exists by (i)[7]) Let $(\vec{v}_1, \ldots, \vec{v}_m)$ be another basis of V. We have $m \geq n$ by definition of the dimension, and we want to show that $m = n$. Let $W = \mathrm{Span}(\vec{e}_2, \ldots \vec{e}_n)$. We have $\dim(W) \leq n - 1$

[7] Choose a generating family $(e_1, \ldots, e_n)$ with minimal cardinality, which is $n = \dim(V)$. If it were not free, we could use the process of the proof of (i) to extract a smaller generating family from it, which would contradict the minimality of the cardinality.

by the induction hypothesis (because $\dim(W) \leq n-1$ and $(\vec{e}_2, \ldots, \vec{e}_n)$ is a basis of W), so $W \neq V$. As $(\vec{v}_1, \ldots, \vec{v}_m)$ generates V, one of the $\vec{v}_i$ has to be in $V - W$. After renumbering, we may assume that it's $\vec{v}_1$. Let's write each $\vec{v}_i$ in the basis $(\vec{e}_1, \ldots, \vec{e}_n)$, $\vec{v}_i = \sum_{j=1}^{n} x_{i,j}\vec{e}_i$. As $\vec{v}_1 \notin W$, $x_{1,1} \neq 0$. For every $i \in \{2, \ldots, m\}$, let $\vec{f_i} = \vec{v}_i - (x_{1,1}^{-1}x_{i,1})\vec{v}_1$. Note that

$$\vec{f_i} = \sum_{j=1}^{n} x_{i,j}\vec{e}_j - \frac{x_{i,1}}{x_{1,1}} \sum_{j=1}^{m} x_{1,j}\vec{e}_j = \sum_{j=2}^{m} (x_{i,j} - \frac{x_{i,1}}{x_{1,1}}x_{1,j})\vec{e}_j,$$

so $\vec{f_i} \in W$. Suppose that we can show that $(\vec{f_2}, \ldots, \vec{f_m})$ is a basis of W. Then the induction hypothesis gives that $m - 1 = \dim(W) = n - 1$, and we can conclude that $n = m$.

So we just need to show that $(\vec{f_2}, \ldots, \vec{f_m})$ is a basis of W. There are two conditions to check, freeness and generation :

- Let $\lambda_2, \ldots, \lambda_m \in K$ such that $\lambda_2\vec{f_2} + \cdots + \lambda_m\vec{f_m} = 0$. Replacing the $\vec{f_i}$ by their values gives

$$\left(-\sum_{i=2}^{m} \lambda_i \frac{x_{i,1}}{x_{1,1}}\right)\vec{v}_1 + \sum_{i=2}^{m} \lambda_i\vec{v}_i = 0.$$

  As $(\vec{v}_1, \ldots, \vec{v}_m)$ is free, this implies that $\lambda_2 = \cdots = \lambda_m = 0$.

- Let $\vec{v} \in W$. As $(\vec{v}_1, \ldots, \vec{m})$ spans V (and $\vec{v}$ is also an element of W), we can write $\vec{v} = \sum_{i=1}^{m} \lambda_i\vec{v}_i$. By the definition of the $\vec{f_i}$,

$$\vec{v} = \sum_{i=1}^{m} \lambda_i\vec{v}_i = \sum_{i=2}^{m} \lambda_i\vec{f_i} + \left(\lambda_1 - \sum_{i=2}^{m} \lambda_i \frac{x_{i,1}}{x_{1,1}}\right)\vec{v}_1.$$

  As $\vec{v}, \vec{f_2}, \vec{f_m} \in W$, this implies that $\left(\lambda_1 - \sum_{i=2}^{m} \lambda_i \frac{x_{i,1}}{x_{1,1}}\right)\vec{v}_1 \in W$. But $\vec{v}_1 \notin W$, so this forces $\lambda_1 - \sum_{i=2}^{m} \lambda_i \frac{x_{i,1}}{x_{1,1}} = 0$, and hence $\vec{v} = \sum_{i=2}^{m} \lambda_i\vec{f_i}$. We have shown that $(\vec{f_2}, \ldots, \vec{f_m})$ spans W.

□

21/2/2017

By (i) and (ii) of theorem 9.3 above, we now know that a finite-dimensional vector space V always has a basis, and that the dimension of V is just equal to the cardinality of any basis. So we can give examples.

Example 9.4 (1) The K-vector space K^n is finite-dimensional, and its dimension is n (because that's the cardinality of the canonical basis).

(2) The $\mathbb{C}$-vector space $\mathbb{C}$ is 1-dimensional. But if we see $\mathbb{C}$ as a $\mathbb{R}$-vector, it's 2-dimensional (with basis $(1, i)$ for example). So the dimension depends on the field of scalars K. If we want to make K explicit, we write $\dim_K$ instead of dim, as in $\dim_{\mathbb{C}}(\mathbb{C}) = 1$ and $\dim_{\mathbb{R}}(\mathbb{C}) = 2$.

Another example of this phenomenon : As a $\mathbb{R}$-vector space, $\mathbb{R}$ has dimension 1. But as a $\mathbb{Q}$-vector space, $\mathbb{R}$ is infinite-dimensional.

(3) Suppose that the K-vector V is finite as a set (if $V \neq \{0\}$, this implies that K is a finite field). Then it is automatically finite-dimensional (because it's generated by the finite family of all its vectors).

(4) If I is an infinite set, both K^I and $K^{(I)}$ are infinite-dimensional. (We know this for $K^{(I)}$ because we constructed a particular infinite basis of it in example 8.3. The case of K^I follows because it contains $K^{(I)}$ as a subspace.) In particular, $K[X]$ is infinite-dimensional.

(5) The $\mathbb{R}$-vector space of infinitely differentiable functions from $\mathbb{R}$ to $\mathbb{R}$ is infinite-dimensional. One way to show this is to show that the family of functions $(e^{at})_{a\in\mathbb{R}}$ is free in this space.

Theorem 9.5 *(continued) Let V be a finite-dimensional K-vector space, and let $n = \dim(V)$.*

(iii) If $(\vec{v}_1, \dots, \vec{v}_r)$ is a free family of V and $(\vec{w}_1, \dots, \vec{w}_s)$ is a generating family of V, then there exist $i_1, \dots, i_m \in \{1, \dots, s\}$ such that $(\vec{v}_1, \dots, \vec{v}_r, \vec{w}_{i_1}, \dots, \vec{w}_{i_m})$ is a basis of V. (This is sometimes called the incomplete basis theorem.)

(iv) Any generating family of V has cardinality $\geq n$, and any free family has cardinality $\leq n$.

(v) Let $(\vec{v}_1, \dots, \vec{v}_n)$ be a family of V (note that the cardinality of the family is $\dim(V)$). Then :

$$(\vec{v}_1, \dots, \vec{v}_n) \text{ is free } \Leftrightarrow (\vec{v}_1, \dots, \vec{v}_n) \text{ generates } V \Leftrightarrow (\vec{v}_1, \dots, \vec{v}_n) \text{ is a basis}$$

(vi) If W is a K-subspace of V, then W is finite-dimensional, and $\dim(W) \leq \dim(V)$, and this is an equality if and only if $W = V$.

Proof.

(iii) Let X be the set of subsets I of $\{1, \dots, m\}$ such that the family $(\vec{v}_1, \dots, \vec{v}_r, \vec{w}_i, i \in I)$ is free. Then $X \neq \varnothing$ because $\varnothing \in X$. Let J be an element of X that is maximal for the inclusion. Then the family $(\vec{v}_1, \dots, \vec{v}_r, w_j, j \in J)$ is free by hypothesis, and we call W the subspace of V that it spans. If $i \notin J$, then the family $(\vec{v}_1, \dots, \vec{v}_r, \vec{w}_j, j \in J, \vec{w}_i)$ is not free (by maximality of J), so $\vec{w}_i \in W$ (otherwise, by (iii) of proposition 7.5, the family we just wrote would be automatically free). So W contains all the $\vec{w}_j$

for $j \in J$ and for $j \in \{1, \ldots, m\} - J$, that is, it contains all the $\vec{w}_j$ for $j \in \{1, \ldots, m\}$. As $(\vec{w}_1, \ldots, \vec{w}_m)$ spans V, this implies that $W = V$, and so $(\vec{v}_1, \ldots, \vec{v}_r, \vec{w}_j, j \in J)$ is a basis of V.

(iv) Let $(\vec{v}_1, \ldots, \vec{v}_r)$ be a free family of V, and $(\vec{w}_1, \ldots, \vec{w}_s)$ be a generating fanmily of V. By (iii), there exists $i_1, \ldots, i_m \in \{1, \ldots, s\}$ (pairwise distinct) such that $(\vec{v}_1, \ldots, \vec{v}_r, \vec{w}_{i_1}, \ldots, \vec{w}_{i_m})$ is a basis of V, and so $r + m = n$. In particular, $r \leq n$. If we apply this to the case where the family $(\vec{v}_1, \ldots, \vec{v}_r)$ is empty ($r = 0$), we get $m = n$, and so $s \geq m = n$.

(v) If $(\vec{v}_1, \ldots, \vec{v}_n)$ is free but not a basis, then by (iii) we can add vectors to it to make it a basis, but then we get a basis of cardinality $> n$, which contradicts (ii). Similarly, if $(\vec{v}_1, \ldots, \vec{v}_n)$ spans V but is not free, then by (i) we can extract a basis of V of cardinality $< n$ from it, which contradicts (ii).

(vi) A family that is free in W is also free in V. By (v), every free family in W has cardinality $\leq n$, so by proposition 9.6 below W is finite-dimensional, and by (ii) $\dim(W) \leq \dim(V)$. Suppose that $\dim(W) = \dim(V)$. Then a basis of W is a free family of cardinality n in V, hence a basis of V by (v), and so $W = V$.

□

Proposition 9.6 *If V is an infinite-dimensional K-vector space, then it has an infinite free family.*

Proof. We construct the elements $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n, \ldots$ of the infinite free family by induction on n.

First, $V \neq \{0\}$ because $\{0\}$ is finite-dimensional, so we can find a nonzero $\vec{v}_1$ is W. Now assume that $n \geq 1$, and that we have constructed $\vec{v}_1, \ldots, \vec{v}_n$ forming a free family in V. As V is infinite-dimensional, $V \neq \text{Span}(\vec{v}_1, \ldots, \vec{v}_n)$, so we can find $\vec{v}_{n+1} \in V - \text{Span}(\vec{v}_1, \ldots, \vec{v}_n)$. By a lemma above, the family $(\vec{v}_1, \ldots, \vec{v}_{n+1})$ is also free.

□

Remark 9.7 By (vi) of the theorem, every subspace of K^n is of dimension $\leq n$ (in particular finite-dimensional), and the only dimension n subspace of K^n is K^n itself.

10 Operations on subspaces

Let V be a K-vector space, and let $W_1, \ldots, W_n$ be subspaces of V. Here are two ways to get more subspaces from $W_1, \ldots, W)n$:

(1) The intersection $W_1 \cap \cdots \cap W_n$ is also a subspace of V. Indeed :

- 0 is in every W_i, so it's in $W_1 \cap \cdots \cap W_n$.

– Let $\vec{v}, \vec{w} \in W_1 \cap \cdots \cap W_n$ and $\lambda \in K$. For every $i \in \{1, \ldots, n\}$, $\vec{v} + \vec{w}$ and $\lambda\vec{v}$ are in W_i because W_i is a subspace. So $\vec{v} + \vec{w}$ and $\lambda\vec{v}$ are in $W_1 \cap \cdots \cap W_n$.

(2) The *sum* $W_1 + \cdots + W_n$ is by definition the subset of V of elements of the form $\vec{v}_1 + \cdots + \vec{v}_n$, with $\vec{v}_1 \in W_1$,…,$\vec{v}_n \in W_n$. This is also a subspace of V. Indeed :

– $0 = 0 + \cdots + 0$ with $0 \in W_i$, so $0 \in W_1 + \cdots + W_n$.

– Let $\vec{v}, \vec{w} \in W_1 + \cdots + W_n$ and $\lambda \in K$. Write $\vec{v} = \vec{v}_1 + \cdots + \vec{v}_n$ and $\vec{w} = \vec{w}_1 + \cdots + \vec{w}_n$ with $\vec{v}_i, \vec{w}_i \in W_i$. Then

$$\vec{v} + \vec{w} = (\vec{v}_1 + \vec{w}_1) + \cdots + (\vec{v}_n + \vec{w}_n) \in W_1 + \cdots + W_n$$

and

$$\lambda\vec{v} = \lambda\vec{v}_1 + \cdots + \lambda\vec{v}_n \in W_1 + \cdots + W_n.$$

Exercise 10.1 (1) In $\mathbb{R}^2$ or $\mathbb{R}^3$, determine the intersection and sum of two lines, a line and a plane, or two planes.

(2) These definitions generalize to infinite families. How ?

Definition 10.2 Let $V, W_1, \ldots, W_n$ be as above, and let $W = W_1 + \cdots + W_n$. We say that the sum of the W_i is *direct* (or that the W_i are in *direct sum*, or that W is the *direct sum* of the W_i) and write $W = W_1 \oplus \cdots \oplus W_n$ if, for every $\vec{v} \in W$, there exist a *unique* family $(\vec{v}_1, \ldots, \vec{v}_n)$ such that $\vec{v}_i \in W_i$ for every i and that $\vec{v} = \vec{v}_1 + \cdots + \vec{v}_n$.

Lemma 10.3 *Let $W_1, \ldots, W_n$ be subspaces of V. Then they are in direct sum if and only if for every $\vec{v}_1 \in W_1, \ldots, \vec{v}_n \in W_n$,*

$$\vec{v}_1 + \cdots + \vec{v}_n = 0 \Rightarrow \vec{v}_1 = \cdots = \vec{v}_n = 0.$$

Proof. If $W_1, \ldots, W_n$ are in direct sum, the condition is clearly true (because we can write $0 = 0 + \cdots + 0$ with $0 \in W_1$,…, $0 \in W_n$, and this must be the unique way).

So suppose that the second condition is true, and let's show that $W_1, \ldots, W_n$ are in direct sum. Let $\vec{v} \in W_1 + \cdots + W_n$, and suppose that we haev $\vec{v} = \vec{w}_1 + \cdots + \vec{w}_n = \vec{w}'_1 + \cdots + \vec{w}'_n$, with $\vec{w}_i, \vec{w}'_i \in W_i$ for every i. We want to show that $\vec{w}_i = \vec{w}'_i$ for every i. But we have

$$(\vec{w}_1 - \vec{w}'_1) + \cdots + (\vec{w}_n - \vec{w}'_n) = 0,$$

with $\vec{w}_i - \vec{w}'_i \in W_i$, so $\vec{w}_1 - \vec{w}'_1 = \cdots = \vec{w}_n - \vec{w}'_n = 0$.

□

Now let's see the relation of sums with bases and dimension.

Proposition 10.4 *Let V be a K-vector space and W_1, W_2 be finite-dimensional K-subspaces of V such that $W_1 + W_2 = V$. Then :*

(i) V is also finite-dimensional, and we have $\dim(V) \leq \dim(W_1) + \dim(W_2)$.

(ii) We have $\dim(V) = \dim(W_1) + \dim(W_2)$ *if and only the sum* $W_1 + W_2$ *is direct (that it, if and only* $V = W_1 \oplus W_2$*).*

(iii) Suppose that $V = W_1 \oplus W_2$*, that* $(\vec{e}_1, \dots, \vec{e}_r)$ *is a basis of* W_1 *and that* $(\vec{f}_1, \dots, \vec{f}_s)$ *is a basis of* W_2*. Then* $(\vec{e}_1, \dots, \vec{e}_r, \vec{f}_1, \dots, \vec{f}_s)$ *is a basis of* V*.*

This generalizes to more than two subspaces (it's an easy induction).

Proof. Let $(\vec{e}_1, \dots, \vec{e}_r)$ be a basis of W_1 and $(\vec{f}_1, \dots, \vec{f}_s)$ be a basis of W_2. Then the family $(\vec{e}_1, \dots, \vec{e}_r, \vec{f}_1, \dots, \vec{f}_s)$ generates $W_1 + W_2 = V$, so V is finite-dimensional and $\dim(V) \leq \dim(W_1) + \dim(W_2)$. This proves (i).

Suppose that $V = W_1 \oplus W_2$, and let's show that $(\vec{e}_1, \dots, \vec{e}_r, \vec{f}_1, \dots, \vec{f}_s)$ is a basis of V, which will show (iii) and half of (ii). We already know that this family is generating, so we only need to show that it's free. Let $\lambda_1, \dots, \lambda_r, \mu_1, \dots, \mu_s \in K$ such that $\lambda_1\vec{e}_1 + \dots + \lambda_r\vec{e}_r + \mu_1\vec{f}_1 + \dots + \mu_s\vec{f}_s = 0$. By lemma 10.3, $\lambda_1\vec{e}_1 + \dots + \lambda_r\vec{e}_r = \mu_1\vec{f}_1 + \dots + \mu_s\vec{f}_s = 0$. As the families $(\vec{e}_1, \dots, \vec{e}_r)$ and $(\vec{f}_1, \dots, \vec{f}_s)$ are both free, this implies that $\lambda_1 = \dots = \lambda_r = \mu_1 = \dots = \mu_s = 0$.

Suppose that $\dim(V) = \dim(W_1) + \dim(W_2) = r + s$. Then by (vi) of theorem 9.3, $(\vec{e}_1, \dots, \vec{e}_r, \vec{f}_1, \dots, \vec{f}_s)$ is a basis of V, and so in particular it's free. Let's show that $V = W_1 \oplus W_2$. We use lemma 10.3 again. So let $\vec{v}_1 \in W_1$ and $\vec{v}_2 \in W_2$ be such taht $\vec{v}_1 + \vec{v}_2 = 0$. We can write $\vec{v}_1 = \lambda_1\vec{e}_1 + \dots + \lambda_r\vec{e}_r$ and $\vec{v}_2 = \mu_1\vec{f}_1 + \dots + \mu_s\vec{f}_s$, with $\lambda_1, \dots, \lambda_r, \mu_1, \dots, \mu_s \in K$. Then $\lambda_1\vec{e}_1 + \dots + \lambda_r\vec{e}_r + \mu_1\vec{f}_1 + \dots + \mu_s\vec{f}_s = \vec{v}_1 + \vec{v}_2 = 0$, so all the λ_i and all the μ_j have to be 0, and this gives that $\vec{v}_1 = \vec{v}_2 = 0$.

□

Theorem 10.5 *Let* V *be a finite-dimensional* K*-vector space, and let* W *be a* K*-subspace of* V*. Then there exists another* K*-subspace* W' *of* V *such that* $V = W \oplus W'$*.*

Proof. Let $(\vec{e}_1, \dots, \vec{e}_r)$ be a basis of W. Then the family $(\vec{e}_1, \dots, \vec{e}_r)$ is free in V, so by (iii) theorem 9.3, we can find $\vec{e}_{r+1}, \dots, \vec{e}_n \in V$ such that $(\vec{e}_1, \dots, \vec{e}_n)$ is a basis of V. Let $W' = \mathrm{Span}(\vec{e}_{r+1}, \dots, \vec{e}_n)$. I claim that $V = W + W'$.

Indeed, if $\vec{v} \in V$, then we can write $\vec{v} = \lambda_1\vec{e}_1 + \dots + \lambda_n\vec{e}_n$, with $\lambda_1, \dots, \lambda_n \in K$. Setting $\vec{w} = \lambda_1\vec{e}_1 + \dots + \lambda_r\vec{e}_r$ and $\vec{w}' = \lambda_{r+1}\vec{e}_{r+1} + \dots + \lambda_n\vec{e}_n$, we see that $\vec{w} \in W$, $\vec{w}' \in W'$ and $\vec{v} = \vec{w} + \vec{w}'$. This shows that $V = W + W'$. But we also know that $\dim(V) = n = r + (n - r) = \dim(W) + \dim(W')$, so $V = W \oplus W'$ by (ii) of proposition 10.4.

□

Corollary 10.6 *Let* V *be a* K*-vector space and* W_1, W_2 *be two subspaces of* V*. Then*

$$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2).$$

Proof. By theorem 10.5, there exist subspaces W_1' of W_1 and W_2' of W_2 such that $W_1 = (W_1 \cap W_2) \oplus W_1'$ and $W_2 = (W_1 \cap W_2) \oplus W_2'$. I claim that $W_1 + W_2 = (W_1 \cap W_2) \oplus W_1' \oplus W_2'$.

This will suffice, by proposition 10.4, we will then have

$$\dim(W_1) = \dim(W_1 \cap W_2) + \dim(W_1')$$

$$\dim(W_2) = \dim(W_1 \cap W_2) + \dim(W_2')$$

and

$$\dim(W_1 + W_2) = \dim(W_1 \cap W_2) + \dim(W_1') + \dim(W_2'),$$

which easily implies the equality we want to prove.

Now let's prove the claim. Let $\vec{v} \in W_1 + W_2$. By definition of $W_1 + W_2$, we have $\vec{v} = \vec{w}_1 + \vec{w}_2$, with $\vec{w}_1 \in W_1$ and $\vec{w}_2 \in W_2$. Also, we can write $\vec{w}_1 = \vec{w}_1' + \vec{y}_1$ and $\vec{w}_2 = \vec{w}_2' + \vec{y}_2$, with $\vec{w}_1' \in W_1'$, $\vec{w}_2' \in W_2'$ and $\vec{y}_1, \vec{y}_2 \in W_1 \cap W_2$. Then $\vec{v} = (\vec{y}_1 + \vec{y}_2) + \vec{w}_1' + \vec{w}_2' \in (W_1 \cap W_2) + W_1' + W_2'$. This proves that $W_1 + W_2 = (W_1 \cap W_2) + W_1' + W_2'$.

Let's prove that the sum is direct. Let $\vec{w}_1 \in W_1'$, $\vec{w} \in W_1 \cap W_2$ and $\vec{w}_2 \in W_2'$ such that $\vec{w}_1 + \vec{w} + \vec{w}_2 = 0$. Then we have $\vec{w}_2 = -(\vec{w}_1 + \vec{w}) \in W_2 \cap W_1$ (because the left hand side is in W_2 and the right hand side is in W_1). As W_2' and $W_1 \cap W_2$ are in direct sum, this forces $\vec{w}_2 = 0$ (otherwise we'd have another decomposition of 0 : $0 = \vec{w}_2 + (-\vec{w}_2)$, with $\vec{w}_2 \in W_2'$ and $-\vec{w}_2 \in W_1 \cap W_2$). So $\vec{w} + \vec{w}_1 = 0$. As $W_1 \cap W_2$ and W_1' are in direct sum, this implies that $\vec{w} = \vec{w}_1 = 0$.

□

23/2/2017

11 Image and kernel of a matrix

Remember the following definitions (definition 4.3) :

Definition 11.1 Let $A \in M_{nm}(K)$. The *image* $\mathrm{Im}(A)$ is the subspace of K^n spanned of the column vectors of A, and the *kernel* $\mathrm{Ker}(A)$ of A is the subset of $\vec{v} \in K^m$ such that $A\vec{v} = 0$.

We have already seen that $\mathrm{Im}(A)$ is a subspace of K^n. Let's show that $\mathrm{Ker}(A)$ is a subspace of K^m :

Lemma 11.2 *For every $A \in M_{nm}(K)$, $\mathrm{Ker}(A)$ is a subspace of K^m.*

Proof. We have $0 \in \mathrm{Ker}(A)$ because $A0 = 0$. If $\vec{v}, \vec{w} \in \mathrm{Ker}(A)$ and $\lambda \in K$, we have $A(\vec{v} + \vec{w}) = A\vec{v} + A\vec{w} = 0 + 0 = 0$, so $\vec{v} + \vec{w} \in \mathrm{Ker}(A)$, and $A(\lambda\vec{v}) = \lambda(A\vec{v}) = \lambda 0 = 0$ [8], so $\lambda\vec{v} \in K$.

□

[8] Exercise : $A(\lambda B) = (\lambda A)B = \lambda(AB)$ if $A \in M_{nm}(K)$, $B \in M_{mp}(K)$ and $\lambda \in K$

Remark 11.3 Let $A \in M_{nm}(K)$, and let $(\vec{e}_1, \ldots, \vec{e}_m)$ be the canonical basis of K^m. Then $A\vec{e}_i$ is the ith column of A for every i, so $\mathrm{Im}(A) = \mathrm{Span}(A\vec{e}_1, \ldots, A\vec{e}_m)$. As multiplication by A sends linear combinations to linear combinations and $(\vec{e}_1, \ldots, \vec{e}_m)$ is a basis of K^m, this means that $\mathrm{Im}(A) = \{A\vec{v}, \vec{v} \in K^m\}$.

Lemma 11.4 *Let A be an invertible $n \times n$ matrix.*

(i) Let $\vec{v}_1, \ldots, \vec{v}_r$ be vectors in K^n. Then the family $(\vec{v}_1, \ldots, \vec{v}_r)$ is free if and only if $(A\vec{v}_1, \ldots, A\vec{v}_r)$ is free.

(ii) If W is a subspace of K^n, $AV := \{A\vec{v}, \vec{v} \in V\}$ is also a subspace, and then $\dim(W) = \dim(A(W))$, where $A(W) = \{A\vec{v}, \vec{v} \in V\}$. In fact, if $(\vec{v}_1, \ldots, \vec{v}_r)$ is a basis of W, then $(A\vec{v}_1, \ldots, A\vec{v}_r)$ is a basis of $A(W)$.

Proof. Let B be the inverse of A.

Let's prove (i). Assume that $(\vec{v}_1, \ldots, \vec{v}_r)$ is free. Let $\lambda_1, \ldots, \lambda_r \in K$ be such that $\lambda_1(A\vec{v}_1) + \ldots \lambda_r(A\vec{v}_r) = 0$. Multiplying by B on the left gives $\lambda_1\vec{v}_1 + \cdots + \lambda_r\vec{v}_r = 0$, which implies $\lambda_1 = \cdots = \lambda_r = 0$. So $(A\vec{v}_1, \ldots, A\vec{v}_r)$ is also free. This gives one direction of (i). The other direction follows from what we just proved, applied to B instead of A (because $(\vec{v}_1, \ldots, \vec{v}_r) = (BA\vec{v}_1, \ldots, BA\vec{v}_r)$.)

Let's prove (ii). The fact that AV is also a subspace follows from the properties of matrix multiplication (see lemma 3.5). Let $(\vec{v}_1, \ldots, \vec{v}_r)$ be a basis of V. Then by (i) the family $(A\vec{v}_1, \ldots, A\vec{v}_r)$ of AV is free, so by (iv) of theorem 9.3 we have $\dim(V) = r \leq \dim(AV)$. As $V = B(AV)$ and B is invertible too, we also get $\dim(AV) \leq \dim(V)$, and this gives (ii).

□

Theorem 11.5 *Let A be a $n \times m$ matrix.*

(A) If A is in reduced row echelon form, then $\mathrm{rank}(A) = \dim(\mathrm{Im}(A)) = \dim(\mathrm{Im}(A^T))$.

(B) If B is a matrix in reduced row echelon form obtained after applying a finite number of elementary row operations to A (which is always possible by theorem 5.4), then $\dim(\mathrm{Im}(A)) = \dim(\mathrm{Im}(B))$ and $\dim(\mathrm{Im}(B^T)) = \dim(\mathrm{Im}(A^T))$.

Remember that $\mathrm{rank}(A)$ is the number of pivots if A is in reduced row echelon form. On the other hand, $\mathrm{Im}(A)$ is the span of the columns of A, a subspace of K^n. As for $\mathrm{Im}(A^T)$, it's the span of the column vectors of A^T, which is the same as the span of the row vectors of A, and is a subspace of K^m. The theorem says in particular that the number of pivots of a reduced row echelon form of A only depends on A, and is equal to $\dim(\mathrm{Im}(A))$, so we may set $\mathrm{rank}(A) = \dim(\mathrm{Im}(A))$ (which is the usual definition). The theorem then also gives that $\mathrm{rank}(A) = \mathrm{rank}(A^T)$.

Proof. Let's prove (A). Suppose that A is in reduced row echelon form. Let $r = \mathrm{rank}(A)$, and write $\vec{l}_1, \ldots, \vec{l}_n$ for the rows of A. Each row either contains a pivot or contains only

zeros, and the rows with only zeros must me at the bottom. So $\vec{l}_{r+1} = \cdots = \vec{l}_n = 0$, and $\mathrm{Im}(A^T) = \mathrm{Span}(\vec{l}_1, \ldots, \vec{l}_r)$ has dimension $\leq r$. To show that $\mathrm{rank}(A) = \dim(\mathrm{Im}(A^T))$, we have to show that the family $(\vec{l}_1, \ldots, \vec{l}_r)$ is free. Write $\vec{l}_i = (a_{i,1}, \ldots, a_{i,m})$. Because A is in reduced row echelon form, we have integers $1 \leq j_1 < j_2 < \cdots < j_r \leq m$ such that, for every $i \in \{1, \ldots, r\}$:

- $a_{i,j} = 0$ for $i < j_i$;
- $a_{i,j_i} = 1$.

Let $x_1, \ldots, x_r \in K$ such that $x_1\vec{l}_1 + \cdots + x_r\vec{l}_r = 0$, and suppose that not all the x_i are zero. Let i be the smallest index such that $x_i \neq 0$. Then the j_ith coefficient of $x_1\vec{l}_1 + \cdots + x_r\vec{l}_r$ is $0 = \sum_{s=1}^{n} x_s a_{s,j_i} = x_i a_{i,j_i} = x_i$, which is a contradiction. So the family $(\vec{l}_1, \ldots, \vec{l}_r)$ is free, and we have shown that $\mathrm{rank}(A) = \dim(\mathrm{Im}(A^T))$.

We still have to show that $\mathrm{rank}(A) = \dim(\mathrm{Im}(A))$. Let $\vec{c}_1, \ldots, \vec{c}_m \in K^n$ be the columns of A, and let V be the subspace of K^n formed of vectors $(x_1, \ldots, x_n)$ such that $x_{r+1} = \cdots = x_n = 0$. Then a basis of V is given by the first r elements of the canonical basis $(\vec{e}_1, \ldots, \vec{e}_n)$ of K^n (defined in example 8.3), so $\dim(V) = r$. As rows $r+1$ to n of A are 0, the columns of A are all in V, and so $\mathrm{Im}(A) \subset V$, and $\dim(\mathrm{Im}(A)) \leq r$ (by (vi) of theorem 9.3). To finish the proof that $r = \dim(\mathrm{Im}(A))$, we have to show that $\mathrm{Im}(A) = V$, that is, that the columns $\vec{c}_1, \ldots, \vec{c}_m$ span V. Remember the integers $1 \leq j_1 < j_2 < \cdots < j_r \leq m$ defined above. The property defining them translates to : $c_{j_i} = (*, \ldots, *, 1, 0, \ldots, 0)$, where the 1 comes in the ith position. Let's show that $\vec{e}_i$ is in $\mathrm{Span}(\vec{c}_{j_1}, \ldots, \vec{c}_{j_i})$ for every $i \in \{1, \ldots, r\}$, which will finish the proof. We do an induction on i. The result is clear for $i = 1$, because $\vec{e}_1 = \vec{c}_{j_1}$. Suppose that $i \geq 2$, and that the result is known for $1, 2, \ldots, i-1$. Then $e_i - \vec{c}_{j_i} \in \mathrm{Span}(\vec{e}_1, \ldots, \vec{e}_{i-1}) \subset \mathrm{Span}(\vec{c}_{j_1}, \ldots, \vec{c}_{j_{i-1}})$ (by the induction hypothesis), and so we indeed have $\vec{e}_i \in \mathrm{Span}(\vec{c}_{j_1}, \ldots, \vec{c}_{j_i})$.

Now let's show (B). We just need to show that $\dim(\mathrm{Im}(A))$ and $\dim(\mathrm{Im}(A^T))$ don't change if we perform one elementary row operation on A. So let's perform an elementary row operation on A, which corresponds to multiplying A on the left by some invertible $C \in M_n(K)$ as explained in definition 5.3. Let $B = CA$. We see easily that $\mathrm{Im}(B) = C\,\mathrm{Im}(A)$ (where $C\,\mathrm{Im}(A)$ is defined in lemma 11.4, and so $\dim(\mathrm{Im}(B)) = \dim(\mathrm{Im}(A))$ by that same lemma. We also have to show that $\dim(\mathrm{Im}(B^T)) = \dim(\mathrm{Im}(A^T))$. In fact we'll show that $\mathrm{Im}(B^T) = \mathrm{Im}(A^T)$. Let $\vec{l}_1, \ldots, \vec{l}_n \in K^n$ be the rows of A. We have three cases, corresponding to the three types of elementary row operations :

(1) B is obtained from A by switching rows i and j. Then obviously the span of the rows of B is the same as the span of the rows of A (because the span doesn't depend on the order of the vectors).

(2) B is obtained from A by multiplying row i by $a \in K - \{0\}$. Then the rows of B are all in $\mathrm{Span}(\vec{l}_1, \ldots, \vec{l}_n)$, so $\mathrm{Im}(B^T) \subset \mathrm{Im}(A^T)$. As we can get A from B by performing an elementary row operation of the same type (multiplying row i by a^{-1}), the same reasoning shows that $\mathrm{Im}(A^T) \subset \mathrm{Im}(B^T)$.

(3) B is obtained from A by adding a times row i to row j ($a \in K$, $i \neq j$). Then again, every row of B is in $\mathrm{Span}(\vec{l_1}, \ldots, \vec{l_n})$, so $\mathrm{Im}(B^T) \subset \mathrm{Im}(A^T)$. As A can be obtained from B by performing the same type of elementary row operation (adding $-a$ times row i to row j), the same reasoning gives that $\mathrm{Im}(B^T) \subset \mathrm{Im}(A^T)$.

□

Theorem 11.6 *(Rank-nullity theorem.) For every $A \in M_{nm}(K)$,*

$$\mathrm{rank}(A) + \dim(\mathrm{Ker}(A)) = m.$$

In other words,

$$\dim(\mathrm{Im}(A)) + \dim(\mathrm{Ker}(A)) = m.$$

Proof. Let's choose a basis $(\vec{v}_1, \ldots, \vec{v}_r)$ of $\mathrm{Ker}(A)$ and a basis $\vec{w}_1, \ldots, \vec{w}_s$ of $\mathrm{Im}(A)$. By remark 11.3, we can find $\vec{u}_1, \ldots, \vec{u}_s \in K^m$ such that $\vec{w}_j = A\vec{u}_j$ for every $j \in \{1, \ldots, s\}$. Let's show that $(\vec{v}_1, \ldots, \vec{v}_r, \vec{u}_1, \ldots, \vec{u}_s)$ is a basis of K^m, which will imply that $m = r + s = \dim(\mathrm{Ker}(A)) + \dim(\mathrm{Im}(A))$ as desired.

First we show that this family spans K^m. Let $\vec{v} \in K^m$. As $A\vec{v} \in \mathrm{Im}(A)$, we can write $A\vec{v} = \mu_1\vec{w}_1 + \cdots + \mu_s\vec{w}_s$. Let $\vec{w} = \vec{v} - (\mu_1\vec{u}_1 + \cdots + \mu_s\vec{u}_s)$. Then

$$A(\vec{v} - \vec{w}) = (\mu_1\vec{w}_1 + \cdots + \mu_s\vec{w}_s) - (\mu_1\vec{w}_1 + \cdots + \mu_s\vec{w}_s) = 0,$$

so $\vec{v} - \vec{w} \in \mathrm{Ker}(A) = \mathrm{Span}(\vec{v}_1, \ldots, \vec{v}_r)$, so $\vec{v} \in \mathrm{Span}(\vec{v}_1, \ldots, \vec{v}_r) + \mathrm{Span}(\vec{u}_1, \ldots, \vec{u}_s)$, as desired.

Then we show that this family is free. Let $\lambda_1, \ldots, \lambda_r, \mu_1, \ldots, \mu_s \in K$ such that $\lambda_1\vec{v}_1 + \cdots + \lambda_r\vec{v}_r + \mu_1\vec{u}_1 + \cdots + \mu_s\vec{u}_s = 0$. Applying A to this gives $\mu_1\vec{w}_1 + \cdots + \mu_s\vec{w}_s = 0$. As $(\vec{w}_1, \ldots, \vec{w}_s)$ is free, this implies that $\mu_1 = \cdots = \mu_s = 0$. But then $\lambda_1\vec{v}_1 + \cdots + \lambda_r\vec{v}_r = 0$. As $(\vec{v}_1, \ldots, \vec{v}_r)$ is free, this implies that $\lambda_1 = \cdots = \lambda_r = 0$.

□

Theorem 11.7 *Let A and B be a square $n \times n$ matrices. Then*

$$AB = I_n \Leftrightarrow BA = I_n.$$

In particular, to check that B is the inverse of A, we only need to check that $BA = I_n$ (or that $AB = I_n$; but not both, that's the point).

Proof. We only need to prove that $AB = I_n$ implies $BA = I_n$. (We'll then get the reverse implication by exchanging the roles of A and B.) So suppose that $AB = I_n$. Then we have $A(B\vec{v}) = (AB)\vec{v} = \vec{v}$ for every $\vec{v} \in K^n$, so by the description of $\mathrm{Im}(A)$ in remark 11.3, $\mathrm{Im}(A) = K^n$, and so $\mathrm{rank}(A) = n$. Let $(\vec{v}_1, \ldots, \vec{v}_n)$ be the rows of A and $(\vec{e}_1, \ldots, \vec{e}_n)$ be the canonical basis of K^n. (We see all these as rwo vectors.) By the fact that $\mathrm{rank}(A) = \mathrm{rank}(A^T)$ (theorem 11.5) and (v) of theorem 9.3, $(\vec{v}_1, \ldots, \vec{v}_n)$ is a basis of K^n. So we can write the vectors $\vec{e}_1, \ldots, \vec{e}_n$ in this basis : $\vec{e}_j = \sum_{j=1}^m c_{ij}\vec{v}_i$. Let

$C = (c_{ij}) \in M_n(K)$. Then by definition of C, CA is the matrix with rows $\vec{e}_1, \ldots, \vec{e}_n$, that is, $CA = I_n$. Using the associativity of the product of matrices gives :

$$C = C(AB) = (CA)B = B.$$

□

Corollary 11.8 *Let A be a square $n \times n$ matrix. Then the following are equivalent :*

(i) *A is invertible.*

(ii) $\mathrm{rank}(A) = n$.

(iii) $\mathrm{rank}(A^T) = n$.

(iv) $\mathrm{Ker}(A) = \{0\}$.

(v) *The columns of A form a basis of K^n.*

(vi) *The columns of A span K^n.*

(vii) *The columns of A form a free family in K^n.*

(viii) *The rows of A form a basis of K^n.*

(ix) *The rows of A span K^n.*

(x) *The rows of A form a free family in K^n.*

Proof. We know that (v), (vi) and (vii) are equivalent by (v) of theorem 9.3, and (viii), (ix) and (x) are equivalent for the same reason. Theorem 11.5 gives the equivalence of (ii) and (vi), as well as the equivalence of (iii) and (viii), that of (ii) and (iii), and that of (ii) and (xi). The equivalence of (ii) and (iv) follows from the rank-nullity theorem (theorem 11.6). If A is invertible, then $\mathrm{Ker}(A) = \{0\}$ (because if $A\vec{v} = 0$, then $\vec{v} = A^{-1}(A\vec{v}) = 0$), so (i) implies (iv). If (v) is true, then we can as in the proof of theorem 11.7 find a matrix $C \in M_n(K)$ such that $CA = I_n$, and then theorem 11.7 implies that $AC = I_n$, so that A is invertible; so (v) implies (i).

□

2/28/2017

12 The change of basis formula

The problem is the following : Let V be a finite-dimensional K-vector space, let $\mathfrak{A} = (\vec{v}_1, \ldots, \vec{v}_n)$ and $\mathfrak{B} = (\vec{w}_1, \ldots, \vec{w}_n)$ be two bases of V. If $\vec{x} \in V$, what is the relationship between $[\vec{x}]_{\mathfrak{A}}$ and $[\vec{x}]_{\mathfrak{B}}$ (the column vectors of coordinates of $\vec{v}$ in the bases $\mathfrak{A}$ and $\mathfrak{B}$) ?

The answer is very simple, and rather than trying to memorize it, it's better to remember how to rederive it.

Let A be the $n \times n$ matrix whose ith column is the coordinate vector of $\vec{w}_i$ in the basis $(\vec{v}_1, \ldots, \vec{v}_n)$, that is, $[\vec{w}_i]_{\mathfrak{A}}$. (This is called the change of basis matrix.)

Proposition 12.1 *For every $\vec{x} \in V$,*

$$[\vec{x}]_{\mathfrak{A}} = A[\vec{x}]_{\mathfrak{B}}.$$

Proof. To remember whether you should use A or A^{-1} (that is, which basis you should express in the other), just test the above formula on $\vec{x} = \vec{w}_i$. The left hand side is $[\vec{w}_i]_{\mathfrak{A}}$, which is the ith column of A. The right hand side is $A[\vec{w}_i]_{\mathfrak{B}} = A\vec{e}_i$ (where $(\vec{e}_1, \ldots, \vec{e}_n)$ is the canonical basis of K^n); this is also the ith column of A.

This also gives the proof in the general case. Indeed, let $\vec{x} \in V$, and write $\vec{v} = x_1\vec{w}_1 + \cdots + x_n\vec{w}_n$ with $x_1, \ldots, x_n \in K$. Then it is easy to see that $[\vec{x}]_{\mathfrak{A}} = x_1[\vec{w}_1]_{\mathfrak{A}} + \cdots + x_n[\vec{w}_n]_{\mathfrak{A}}$. On the other hand, $[\vec{x}]_{\mathfrak{B}} = \begin{pmatrix} x_1 \\ \ddots \\ x_n \end{pmatrix}$, so $A[\vec{x}]_{\mathfrak{B}}$ is also equal to $x_1[\vec{w}_1]_{\mathfrak{A}} + \cdots + x_n[\vec{w}_n]_{\mathfrak{A}}$. (Because $[\vec{w}_i]_{\mathfrak{A}}$ is the ith column of A.)

□

Corollary 12.2 *Let $\mathfrak{A}$, $\mathfrak{B}$ and A be as in the proposition, and let B be the $n \times n$ matrix with columns $[\vec{v}_1]_{\mathfrak{B}}, \ldots, [\vec{v}_n]_{\mathfrak{B}}$. (That is, the change of basis matrix in the other direction.)*

Then $AB = BA = I_n$. That is, A and B are invertible and $B = A^{-1}$.

Proof. Let's apply the change of basis formula to the vectors $\vec{v}_1, \ldots, \vec{v}_n$. The matrix with columns $[\vec{v}_1]_{\mathfrak{A}}, \ldots, [\vec{v}_n]_{\mathfrak{A}}$ is just the identity matrix I_n. On the other hands, the matrix with columns $[\vec{v}_1]_{\mathfrak{B}}, \ldots, [\vec{v}_n]_{\mathfrak{B}}$ is B by definition. So the change of basis formula gives $AB = I_n$. By theorem 11.7, this implies that $BA = I_n$. (We could also use the change of basis formula in the other direction.)

□

13 Linear transformations

We fix a field K.

Definition 13.1 If V and W are two K-vector spaces, a *(K)-linear transformation* (or *linear map*) from V to W (also called a *morphism* or *homomorphism* of vector spaces) is a map $f : V \to W$ satisfying the following two conditions :

(1) For every $v, v' \in V$, $f(v + v') = f(v) + f(v')$.

(2) For every $v \in V$ and $\lambda \in K$, $f(\lambda v) = \lambda f(v)$.

Note that we automatically have $f(0) = 0$ (because $0 = 0 + 0$, so (1) implies that that $f(0) = 2f(0)$, hence $f(0) = 0$).

We write $\mathrm{Hom}(V, W)$ (or $\mathrm{Hom}_K(V, W)$ if K is not clear from the context) for the set of linear transformations from V to W. If $V = W$, we also write $\mathrm{End}(V)$ (or $\mathrm{End}_K(V)$) instead of $\mathrm{Hom}(V, V)$, and we call linear transformations from V to V *endomophisms* of V.

Here are some examples of linear transformations.

Example 13.2

- The zero map from V to W (it sends every element of V to $0 \in W$).
- The identity map from V to V.
- The trace : $M_n(K) \to K$, $A \mapsto \mathrm{Tr}(A) := \sum_{i=1}^{n} A_{ii}$.
- $M_{nm}(K) \to M_{mn}(K)$, $A \mapsto A^T$.
- $f : \mathbb{R}^3 \to \mathbb{R}$, $(x_1, x_2, x_3) \mapsto 3x_1 - x_2 + 18x_3$.
- Any linear transformation $f : K \to W$ is of the form $x \mapsto x\vec{w}$, for some $\vec{w} \in W$.
- The real and imaginary part are $\mathbb{R}$-linear transformations from $\mathbb{C}$ to $\mathbb{R}$.
- Let p be a prime number. Suppose that $\mathrm{char}(K) = p$, which means that $p = 0$ in K (for example, this is true if $K = \mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}$). Then the map $\mathrm{Frob} : K \to K$, $x \mapsto x^p$ is a $\mathbb{F}_p$-linear transformation (called the *Frobenius map*).
- The maps $P \mapsto P'$ and $P \mapsto P(X + 2)$ are endormorphisms of $K[X]$.
- Let V be the $\mathbb{R}$-vector space of infinitely differentiable functions from $\mathbb{R}$ to $\mathbb{R}$ (or from $\mathbb{R}$ to $\mathbb{C}$). Then the map $f \mapsto f'$ is an endomorphism of V.
- Let V be the $\mathbb{R}$-vector space of continuous functions from $[0, 1]$ to $\mathbb{R}$. Then the map $f \mapsto \int_0^1 f(t)dt$ is a linear transformation from V to $\mathbb{R}$.
- The map $f \mapsto f(5)$ is a linear transformation from $\mathbb{R}^{\mathbb{R}}$ to $\mathbb{R}$.

Here are few non-examples.

Example 13.3

- $x \mapsto x^2$ is not a linear transformation from K to K, unless K has characteristic 2, i.e. $2 = 0$ in K. (In general, maps that involve squares or higher powers or inverses tend not to be linear, but see the example of the Frobenius.)
- $f \mapsto f(0) + 1$ is not a linear transformation from $\mathbb{R}^{\mathbb{R}}$ to $\mathbb{R}$ (it does not send 0 to 0).

Here are a few easy properties of linear transformations. ("Easy" as in "every proof is a straightforward check and will be left as an exercise".)

Lemma 13.4 *Let V, W, U be three K-vector spaces.*

(i) If $f, g \in \mathrm{Hom}(V, W)$ and $\lambda \in K$, then $f + g$ and λf are also in $\mathrm{Hom}(V, W)$. (Where the sum and scalar multiplication are defined by $(f + g)(v) = f(v) + g(v)$ and $(\lambda f)(v) = \lambda f(v)$, for every $v \in V$).

(ii) With the operations of (i), $\mathrm{Hom}(V, W)$ is a K-vector space.

(iii) If $f \in Hom(V, W)$ and $g \in \mathrm{Hom}(W, U)$, then $g \circ f \in \mathrm{Hom}(V, U)$.

(iv) If $f, f' \in \mathrm{Hom}(V, W)$, $g, g' \in \mathrm{Hom}(W, U)$ and $\lambda \in K$, then

$$f \circ (g + g') = f \circ g + f \circ g'$$

$$(f + f') \circ g = f \circ g + f \circ g'$$

$$(\lambda f) \circ g = f \circ (\lambda g) = \lambda (f \circ g).$$

(Remember that $\circ$ has priority over $+$.)

In particular, $\mathrm{End}(V)$, with the additive and the composition, is a ring (noncommutative in general), just like $M_n(K)$. This is not a coincidence, as we will see later. (Also, this is a useless remark. I just like to say "ring". Ring ring ring ring.)

Finally we can make a precise definition of what is means for two K-vector spaces to "look exactly the same".

Definition 13.5 If V and W are two K-vector spaces, an *isomorphism* from V to W is a linear transformation $f : V \to W$ that is also a bijection. We say that two K-vector spaces are *isomorphic* if there exists an isomorphism between them.

Example 13.6

- When I said in the first week of class that "$M_{nm}(K)$ is the same as K^{nm}", what I actually meant is "these spaces are isomorphic". The isomorphism I gave was the one that sends $A \in M_{nm}(K)$ to the element

$$(A_{11}, A_{12}, \ldots, A_{1m}, A_{21}, A_{22}, \ldots, A_{2m}, \ldots, A_{n1}, \ldots, A_{nm})$$

  of K^{nm} (But there are other isomorphisms.) Note this is totally incompatible with matrix multiplication.

- Let V be a n-dimensional vector space, and let $\mathfrak{B} = (\vec{v}_1, \ldots, \vec{v}_n)$ be a basis of V. Remember that, for $\vec{v} \in V$, we write $[\vec{v}]_{\mathfrak{B}} \in K^n$ for the family of coordinates of $\vec{v}$ in the basis $\mathfrak{B}$. Then the map $\vec{v} \mapsto [\vec{v}]_{\mathfrak{B}}$ is an isomorphism from V to K^n. The inverse isomorphism is $(x_1, \ldots, x_n) \mapsto x_1\vec{v}_1 + \cdots + x_n\vec{v}_n$.

  In particular, all finite-dimensional K-vector spaces of the same dimension are isomorphic (since they are all isomorphic to K^n).

Lemma 13.7 *If $f : V \to W$ is an isomorphism, then its inverse $f^{-1} : W \to V$ is also a linear transformation. (So it is also an isomorphism.)*

Proof. Let $\vec{w}, \vec{w}' \in W$ and $\lambda \in K$. Using the fact that f is linear and that $f \circ f^{-1} = \mathrm{id}$, we get :

$$f(f^{-1}(\vec{w} + \vec{w}')) = \vec{w} + \vec{w}' = f(f^{-1}(\vec{w})) + f(f^{-1}(\vec{w}')) = f(f^{-1}(\vec{w}) + f^{-1}(\vec{w}))$$

and

$$f(f^{-1}(\lambda\vec{w})) = \lambda\vec{w} = \lambda(f(f^{-1}(\vec{w}))) = f(\lambda f^{-1}(\vec{w})).$$

As f is bijective, this implies that

$$f^{-1}(\vec{w} + \vec{w}') = f^{-1}(\vec{w}) + f^{-1}(\vec{w}')$$

and

$$f^{-1}(\lambda\vec{w}) = \lambda f^{-1}(\vec{w}),$$

as desired.

□

Lemma 13.8 *Let $f : V \to W$ be an isomorphism and $(\vec{v}_i)_{i\in I}$ be a family of vectors of V. Then :*

(i) $(\vec{v}_i)_{i\in I}$ is free if and only $(f(\vec{v}_i))_{i\in I}$ is free.

(ii) $(\vec{v}_i)_{i\in I}$ generates V if and only $(f(\vec{v}_i))_{i\in I}$ generates W.

(iii) $(\vec{v}_i)_{i\in I}$ is a basis of V if and only if $(f(\vec{v}_i))_{i\in I}$ is a basis of W.

In particular, two isomorphic vector spaces have the same dimension.

Proof. (iii) follows from (i) and (ii) by definition of a basis. As every isomorphism has an inverse that's also an isomorphism, we only need to prove the implications $\Rightarrow$ in (i) and (ii).

We start with (i). Suppose that $(\vec{v}_i)_{i\in I}$ is free, and let $\lambda_1, \ldots, \lambda_n \in K$ and $i_1, \ldots, i_n \in I$ distinct such that $\lambda_1 f(\vec{v}_{i_1}) + \cdots + \lambda_n(\vec{v}_{i_n}) = 0$. By linearity of f, this gives $f(\lambda_1\vec{v}_{i_1} + \cdots + \lambda_n\vec{v}_{i_n}) = 0$, and using the fact that f is bijective (and that $f(0) = 0$), we get $\lambda_1\vec{v}_{i_1} + \cdots + \lambda\vec{v}_{i_n} = 0$. As $(\vec{v}_i)_{i\in I}$ is free, this implies that $\lambda_1 = \cdots = \lambda_n = 0$, as desired.

Now let's prove (ii). Suppose that $(\vec{v}_i)_{i\in I}$ generates V, and let $\vec{w} \in W$. We write $f^{-1}(\vec{w}) = \sum_{i\in I} \lambda_i\vec{v}_i$, where only a finite number of the λ_i are nonzero. Then

$$\vec{w} = f(f^{-1}(\vec{w})) = f(\sum_{i\in I} \lambda_i\vec{v}_i) = \sum_{i\in I} \lambda_i f(\vec{v}_i).$$

□

14 Linear transformations and matrices

Theorem 14.1 *Let V and W be two K-vectors spaces.*

(i) If $(\vec{v}_i)_{i\in I}$ is a generating family of V and $f, g : V \to W$ are two linear transformations such that $f(\vec{v}_i) = g(\vec{v}_i)$ for every $i \in I$, then $f = g$.

(ii) If $(\vec{v}_i)_{i\in I}$ is a basis of V and $(\vec{w}_i)_{i\in I}$ is a family of vectors in W, then there exists one and only one linear transformation $f : V \to W$ such that $f(\vec{v}_i) = \vec{w}_i$ for every $i \in I$.

Proof.

(i) Let $\vec{v} \in V$, and write $\vec{v} = \sum_{i\in I} \lambda_i \vec{v}_i$, where only a finite number of the λ_i are nonzero. Then

$$f(\vec{v}) = f(\sum_{i\in I} \lambda_i \vec{v}_i) = \sum_{i\in I} \lambda_i f(\vec{v}_i) = \sum_{i\in I} \lambda_i g(\vec{v}_i) = g(\sum_{i\in I} \lambda_i \vec{v}_i) = g(\vec{v}).$$

(ii) The uniqueness follows from (i). Now defined f in the following way : If $\vec{v} \in V$, then there exists a unique family $(\lambda_i)_{i\in I}$ of elements of K such that all but a finite number of the λ_i are zero and that $\vec{v} = \sum_{i\in I} \lambda_i \vec{v}_i$. We set

$$f(\vec{v}) = \sum_{i\in I} \lambda_i \vec{w}_i.$$

It is now easy to check that f is linear.

□

3/1/2017

Definition 14.2 Let V and W be finite-dimensional vector spaces., and fix bases $\mathfrak{A} = (\vec{v}_1, \ldots, \vec{v}_n)$ of V and $\mathfrak{B} = (\vec{w}_1, \ldots, \vec{w}_m)$ of W. If $f \in \mathrm{Hom}(V, W)$, we write $[f]_{\mathfrak{B},\mathfrak{A}}$ for the $m \times n$ matrix with columns $[f(\vec{v}_1)]_{\mathfrak{B}}, \ldots, [f(\vec{v}_n)]_{\mathfrak{B}}$ and call it *matrix of f in the bases $\mathfrak{A}$ and $\mathfrak{B}$*.

If $V = W$ and $\mathfrak{A} = \mathfrak{B}$, we'll just write $[f]_{\mathfrak{B}}$ instead of $[f]_{\mathfrak{B},\mathfrak{B}}$.

Corollary 14.3 *Let $V, W, \mathfrak{A}, \mathfrak{B}$ be as the definition above. Then the map*

$$\begin{cases} \mathrm{Hom}(V, W) & \to & M_{m,n}(K) \\ f & \mapsto & [f]_{\mathfrak{B},\mathfrak{A}} \end{cases}$$

is an isomorphism of K-vector spaces. Its inverse is defined as follows : a matrix $A \in M_{mn}(K)$ goes to the linear trasnformation that sends $\vec{v} \in V$ to the unique vector

$\vec{w} \in W$ *such that* $[\vec{w}]_{\mathfrak{B}} = A[\vec{v}]_{\mathfrak{A}}$ *(the right hand side is the matrix product of* A *by the column vector* $[\vec{v}]_{\mathfrak{A}}$*).*

Moreover, if U *is a third vector space with a basis* $\mathfrak{C}$*, then for every* $f \in \mathrm{Hom}(V, W)$ *and* $g \in \mathrm{Hom}(W, U)$*,*

$$[g \circ f]_{\mathfrak{C},\mathfrak{A}} = [g]_{\mathfrak{C},\mathfrak{B}}[f]_{\mathfrak{B},\mathfrak{A}}$$

(the product on the right hand side is the matrix product).

Remark 14.4 Note the very important corollary of the corollary : If V and W are finite-dimensional vector spaces, and if we write $n = \dim V$ and $m = \dim W$, then :

$$\dim(\mathrm{Hom}(V, W)) = \dim(M_{m,m}(K)) = nm = (\dim V)(\dim W).$$

Proof. By (ii) of theorem 14.1 above, the map $\mathrm{Hom}(V, W) \to W^n$, $f \mapsto (f(\vec{v}_1), \ldots, f(\vec{v}_n))$, is bijective. As it is clearly linear, it's an isomorphism. To finish the proof of the first statement, we just need to notice that, by definition of a basis, the map $W \to M_{m1}(K)$, $\vec{w} \mapsto [\vec{w}]_{\mathfrak{B}}$, is an isomorphism.

Let's prove the second sentence of the corollary. Let $f \in \mathrm{Hom}(V, W)$. For every $r \in \{1, \ldots, n\}$, we write $f(\vec{v}_r) = \sum_{s=1}^{m} a_{sr}\vec{w}_s$. Then $[f]_{\mathfrak{B},\mathfrak{A}} = (a_{ij})_{1 \le i \le m, 1 \le j \le n}$. Now let $\vec{v} \in V$, and write $\vec{v} = x_1\vec{v}_1 + \cdots + x_n\vec{v}_n$, that is, $[\vec{v}]_{\mathfrak{A}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. Then the coefficient of $\vec{w}_s$ in $f(\vec{v})$ is $\sum_{r=1}^{n} a_{sr}x_r$, which is also the $(s, 1)$-entry of the $m \times 1$ matrix $[f]_{\mathfrak{B},\mathfrak{A}}[\vec{v}]_{\mathfrak{A}}$. So we see that we indeed recover f from $[f]_{\mathfrak{B},\mathfrak{A}}$ by the formula of the corollary.

Now let's prove the formula for the composition. Write $\mathfrak{C} = (\vec{u}_1, \ldots, \vec{u}_p)$. We also write $[f]_{\mathfrak{B},\mathfrak{A}} = (a_{ij})$ and $[g]_{\mathfrak{C},\mathfrak{B}} = (b_{ij})$, which means that $f(\vec{v}_r) = \sum_{s=1}^{m} a_{sr}\vec{w}_s$ and $g(\vec{w}_s) = \sum_{t=1}^{p} b_{ts}\vec{u}_t$. For $r \in \{1, \ldots, n\}$, the rth column of $[g \circ f]_{\mathfrak{C},\mathfrak{A}}$ is $[(g \circ f)(\vec{v}_r)]_{\mathfrak{C}}$. We have

$$(g \circ f)(\vec{v}_r) = g(\sum_{s=1}^{m} a_{sr}\vec{w}_s) = \sum_{t=1}^{p}\sum_{s=1}^{m} b_{ts}a_{sr}\vec{u}_t = \sum_{t=1}^{p}(\sum_{s=1}^{m} b_{ts}a_{sr})\vec{u}_t.$$

Note that the coefficient of $\vec{u}_t$ is just the entry in position (t, s) of the matrix $[g]_{\mathfrak{C},\mathfrak{B}}[f]_{\mathfrak{B},\mathfrak{A}}$. In other words, the rth column of $[g \circ f]_{\mathfrak{C},\mathfrak{A}}$ is equal to the rth column of $[g]_{\mathfrak{C},\mathfrak{B}}[f]_{\mathfrak{B},\mathfrak{A}}$. □

Remark 14.5 Suppose that $V = W = U$ and $\mathfrak{A} = \mathfrak{B} = \mathfrak{C}$. Note that $[\mathrm{id}_V]_{\mathfrak{B}} = I_n$ (the identity matrix).[9]

Then the corollary implies that, for every $f \in \mathrm{End}(V)$: f is an isomorphism if and only if $[f]_{\mathfrak{B}}$ is invertible.

Moreover, if f is an isomorphism, then $[f^{-1}]_{\mathfrak{B}} = [f]_{\mathfrak{B}}$.

[9] Beware, if $\mathfrak{A}$ and $\mathfrak{B}$ are two *dictinct* bases of V, then $[\mathrm{id}_V]_{\mathfrak{B},\mathfrak{A}}$ is not the identity matrix !

Proof. Suppose that f is an isomorphism. Then

$$I_n = [\mathrm{id}_V]_{\mathfrak{B}} = [f^{-1} \circ f]_{\mathfrak{B}} = [f^{-1}]_{\mathfrak{B}}[f]_{\mathfrak{B}}.$$

By theorem 11.7, this implies that $[f]_{\mathfrak{B}}$ is inversible and that its inverse is $[f^{-1}]_{\mathfrak{B}}$.

Conversely, assume that $[f]_{\mathfrak{B}}$ is inversible, and let A be its inverse. By corollary 14.3, there exists a unique $g \in \mathrm{End}(V)$ such that $[g]_{\mathfrak{B}} = A$. Then we have :

$$[g \circ f]_{\mathfrak{B}} = [g]_{\mathfrak{B}}[f]_{\mathfrak{B}} = A[f]_{\mathfrak{B}} = I_n = [\mathrm{id}_V]$$

and

$$[f \circ g]_{\mathfrak{B}} = [f]_{\mathfrak{B}}A = I_n = [\mathrm{id}_V],$$

so $g \circ f = f \circ g = \mathrm{id}_V$.

□

Remark 14.6 Actually, using the same proof as in the previous remark, we can get a slightly more general result : If V and W are K-vector spaces of the same finite dimension, $\mathfrak{A}$ is a basis of V and $\mathfrak{B}$ is a basis of W, then a linear application $f : V \to W$ is an isomorphism if and only if the matrix $[f]_{\mathfrak{B},\mathfrak{A}}$ is invertible. (Note that this matrix is square because $\dim(V) = \dim(W)$, so it makes sense to ask if it's invertible.)

We also have a change of basis formula for the matrix of a linear transformation. Here, we'll just give the most useful case, which is the case where $V = W$. (Again, it is a very bad idea to try to memorize this without understanding it.)

Proposition 14.7 *Let V be a finite-dimensional K-vector space, let $\mathfrak{A} = (\vec{v}_1, \ldots, \vec{v}_n)$ and $\mathfrak{B} = (\vec{w}_1, \ldots, \vec{w}_n)$ be two bases of V. We write A for the matrix with columns $[\vec{w}_1]_{\mathfrak{A}}, \ldots, [\vec{w}_n]_{\mathfrak{A}}$. (The change of basis matrix from $\mathfrak{B}$ to $\mathfrak{A}$.)*

Then, for every $f \in \mathrm{End}(V)$,

$$[f]_{\mathfrak{B}} = A^{-1}[f]_{\mathfrak{A}}A.$$

Proof. By the change of basis formula (and corollary 12.2), for every $\vec{v} \in V$:

$$[\vec{v}]_{\mathfrak{A}} = A[\vec{v}]_{\mathfrak{B}}$$

and

$$[\vec{v}]_{\mathfrak{B}} = A^{-1}[\vec{v}]_{\mathfrak{A}}.$$

Remember also that $[f]_{\mathfrak{B}}$ is the unique $n \times n$ matrix such that, for every $\vec{v} \in V$, $[f(\vec{v})]_{\mathfrak{B}} = [f]_{\mathfrak{B}}[\vec{v}]_{\mathfrak{B}}$ and $[f]_{\mathfrak{A}}$ is the unique $n \times n$ matrix such that, for every $\vec{v} \in V$, $[f(\vec{v})]_{\mathfrak{A}} = [f]_{\mathfrak{A}}[\vec{v}]_{\mathfrak{A}}$

So we have to prove that $A^{-1}[f]_{\mathfrak{A}}A$ satisfies the property that characterizes $[f]_{\mathfrak{B}}$. Let $\vec{v} \in V$. We have :

$$(A^{-1}[f]_{\mathfrak{A}}A)[\vec{v}]_{\mathfrak{B}} = (A^{-1}[f]_{\mathfrak{A}})(A[\vec{v}]_{\mathfrak{B}}) = A^{-1}[f]_{\mathfrak{A}}[\vec{v}]_{\mathfrak{A}} = A^{-1}[f(\vec{v})]_{\mathfrak{A}} = [f(\vec{v})]_{\mathfrak{B}}.$$

□

Example 14.8 Linear transformations from K^n to K^m

In this subsection, we take $V = K^n$ and $W = K^m$, and we use the canonical basis on both sides. If $f \in \mathrm{Hom}(K^n, K^m)$, the matrix of f in the canonical bases is often just called "the matrix of f". This gives an isomorphism $\mathrm{Hom}(K^n, K^m) \xrightarrow{\sim} M_{mn}(K)$, and composition corresponds to matrix multiplication by this isomorphism. This is the main reason that we defined matrix multiplication that way, and also the reason that it's distributive with respect to matrix addition.

In the other direction, if $A \in M_{m,n}(K)$, then we recover the corresponding $f \in \mathrm{Hom}(K^n, K^m)$ by the formula $f(\vec{v}) = A\vec{v}$. (By corollary 14.3.)

In particular, the identity matrix $I_n \in M_n(K)$ corresponding to the identity map id_{K_n} (hence the name). Also, a matrix $A \in M_n(K)$ is invertible if and only if the endomorphism of K^n that it defines is an isomorphism.

Now let's generalize the notions of kernel, image and rank from matrices to linear transformation.

15 Kernel and image of a linear transformation

Let V and W be K-vector spaces.

Definition 15.1 If $f \in \mathrm{Hom}(V, W)$, then :

(i) The *kernel* of f is the subset $\mathrm{Ker}(f)$ of V defined by $\mathrm{Ker}(f) = \{\vec{v} \in V | f(\vec{v}) = 0\}$.

(ii) The *image* of f is the subset $\mathrm{Im}(f)$ of W defined by $\mathrm{Im}(f) = \{\vec{w} \in W | \exists \vec{v} \in V, f(\vec{v}) = \vec{w}\}$.

(iii) The *rank* of f is $\mathrm{rank}(f) = \dim(\mathrm{Im}(f))$. (It's a nonnegative integer or $+\infty$.)

Example 15.2 If $V = K^n$, $W = K^m$, $f \in \mathrm{Hom}(V, W)$ and $A \in M_{mn}(K)$ is the matrix of f (in the canonical bases), then $\mathrm{Ker}(f) = \mathrm{Ker}(A)$, $\mathrm{Im}(f) = \mathrm{Im}(A)$ and $\mathrm{rank}(f) = \mathrm{rank}(A)$.

More examples.

Lemma 15.3 *Let $f : V \to W$ be a linear transformation. Then* $\mathrm{Ker}(f)$ *is a subspace of V and* $\mathrm{Im}(f)$ *is a subspace of W.*

Proof.

- $\mathrm{Ker}(f)$: We have $0 \in \mathrm{Ker}(f)$ because $f(0) = 0$. Let $\vec{v}, \vec{w} \in \mathrm{Ker}(f)$ and $\lambda \in K$. Then
$$f(\vec{v} + \vec{w}) = f(\vec{v}) + f(\vec{w}) = 0 + 0 = 0$$
$$f(\lambda\vec{v}) = \lambda f(\vec{v}) = \lambda 0 = 0$$
so $\vec{v} + \vec{w}$ and $\lambda\vec{v}$ are also in $\mathrm{Ker}(f)$.

- $\mathrm{Im}(f)$: We have $0 \in \mathrm{Im}(f)$ because $0 = f(0)$. Let $\vec{v}, \vec{w} \in \mathrm{Im}(f)$ and $\lambda \in K$. By definition of $\mathrm{Im}(f)$, we have vectors $\vec{v}', \vec{w}'$ in V such that $\vec{v} = f(\vec{v}')$ and $\vec{w} = f(\vec{w}')$. Then
$$\vec{v} + \vec{w} = f(\vec{v}') + f(\vec{w}') = f(\vec{v}' + \vec{w}') \in \mathrm{Im}(f)$$
and
$$\lambda\vec{v} = \lambda f(\vec{v}') = f(\lambda\vec{v}') \in \mathrm{Im}(f).$$

□

The following result is very important, even though its proof is so short.

Proposition 15.4 *Let $f : V \to W$ be a linear transformation. Then f is injective if and only if* $\mathrm{Ker}(f) = \{0\}$.

Proof. Suppose that f is injective. Then, if $\vec{v} \in \mathrm{Ker}(f)$, we have $f(\vec{v}) = 0 = f(0)$, so $\vec{v} = 0$ by injectivity of f.

Suppose that $\mathrm{Ker}(f) = \{0\}$. Let $\vec{v}, \vec{w} \in V$ be such that $f(\vec{v}) = f(\vec{w})$. Then $f(\vec{v} - \vec{w}) = f(\vec{v}) - f(\vec{w}) = 0$, so $\vec{v} - \vec{w} \in \mathrm{Ker}(f)$, so $\vec{v} - \vec{w} = 0$, so $\vec{v} = \vec{w}$.

□

16 Rank-nullity theorem for linear transformations

Here is the rank-nullity theorem for linear transformations.

Theorem 16.1 *Let V and W be finite-dimensional K-vector spaces, and let $f : V \to W$ be a linear transformation. Then :*
$$\dim(\mathrm{Ker}(f)) + \mathrm{rank}(f) = \dim(V).$$

Proof. We can deduce this theorem from the rank-nullity theorem for $n \times m$ matrices (theorem 11.6), by choosing bases $\mathfrak{A}$ of V and $\mathfrak{B}$ of W and relating the kernel and image of f to the kernel and image of the matrix $[f]_{\mathfrak{B},\mathfrak{A}}$ of f in those bases. Here is the relation : Write $A = [f]_{\mathfrak{B},\mathfrak{A}}$. Then $\vec{v} \in V$ is in $\mathrm{Ker}(f)$ if and only if $[\vec{v}]_{\mathfrak{A}} \in \mathrm{Ker}(A)$, and $\vec{w} \in W$ is in $\mathrm{Im}(f)$ if and only if $[\vec{w}]_{\mathfrak{B}} \in \mathrm{Im}(A)$. (This follows from the formula $[f(\vec{v})]_{\mathfrak{B}} = [f]_{\mathfrak{B},\mathfrak{A}}[\vec{v}]_{\mathfrak{A}}$ of corollary 14.3.)

Or we could just adapt the proof of theorem 11.6 to this case. Let's do this as an exercise : Let's choose a basis $(\vec{v}_1, \ldots, \vec{v}_r)$ of $\mathrm{Ker}(f)$ and a basis $\vec{w}_1, \ldots, \vec{w}_s$ of $\mathrm{Im}(f)$. Choose $\vec{u}_1, \ldots, \vec{u}_s \in V$ such that $\vec{w}_j = f(\vec{u}_j)$ for every $j \in \{1, \ldots, s\}$. Let's show that $(\vec{v}_1, \ldots, \vec{v}_r, \vec{u}_1, \ldots, \vec{u}_s)$ is a basis of V, which will imply that $\dim(V) = r + s = \dim(\mathrm{Ker}(f)) + \dim(\mathrm{Im}(f))$ as desired.

First we show that this family spans V. Let $\vec{v} \in V$. As $f(\vec{v}) \in \mathrm{Im}(f)$, we can write $f(\vec{v}) = \mu_1\vec{w}_1 + \cdots + \mu_s\vec{w}_s$. Let $\vec{w} = \vec{v} - (\mu_1\vec{u}_1 + \cdots + \mu_s\vec{u}_s)$. Then
$$f(\vec{v} - \vec{w}) = (\mu_1\vec{w}_1 + \cdots + \mu_s\vec{w}_s) - (\mu_1\vec{w}_1 + \cdots + \mu_s\vec{w}_s) = 0,$$

so $\vec{v} - \vec{w} \in \mathrm{Ker}(f) = \mathrm{Span}(\vec{v}_1, \ldots, \vec{v}_r)$, so $\vec{v} \in \mathrm{Span}(\vec{v}_1, \ldots, \vec{v}_r) + \mathrm{Span}(\vec{u}_1, \ldots, \vec{u}_s)$, as desired.

Then we show that this family is free. Let $\lambda_1, \ldots, \lambda_r, \mu_1, \ldots, \mu_s \in K$ such that $\lambda_1\vec{v}_1 + \cdots + \lambda_r\vec{v}_r + \mu_1\vec{u}_1 + \cdots + \mu_s\vec{u}_s = 0$. Applying f to this gives $\mu_1\vec{w}_1 + \cdots + \mu_s\vec{w}_s = 0$. As $(\vec{w}_1, \ldots, \vec{w}_s)$ is free, this implies that $\mu_1 = \cdots = \mu_s = 0$. But then $\lambda_1\vec{v}_1 + \cdots + \lambda_r\vec{v}_r = 0$. As $(\vec{v}_1, \ldots, \vec{v}_r)$ is free, this implies that $\lambda_1 = \cdots = \lambda_r = 0$.

□

Corollary 16.2 *With the notation of the theorem, suppose that* $\dim(V) = \dim(W)$. *Then :*

$$f \text{ is injective } \Leftrightarrow f \text{ is surjective } \Leftrightarrow f \text{ is an isomorphism.}$$

Proof. The rank-nullity theorem (and the hypothesis on the dimensions) says that $\dim(W) = \dim(V) = \dim(\mathrm{Ker}(f)) + \dim(\mathrm{Im}(f))$. Also, by proposition 15.4, we know that f is injective if and only if $\mathrm{Ker}(f) = \{0\}$, which is equivalent to $\dim(\mathrm{Ker}(f)) = 0$.

Suppose that f is injective. Then $\dim(\mathrm{Ker}(f)) = 0$, so $\dim(\mathrm{Im}(f)) = \dim(W)$. As $\mathrm{Im}(f)$ is a subspace of W, this implies that $W = \mathrm{Im}(f)$ (by (vi) of theorem 9.3).

Suppose that f is surjective, that is, that $W = \mathrm{Im}(f)$. Then $\dim(\mathrm{Ker}(f)) = \dim(W) - \dim(\mathrm{Im}(f)) = 0$, so f is also injective.

□

Remark 16.3 We can also use the rank-nullity theorem to prove that, if W_1 and W_2 are two finite-dimensional subspaces of a vector space W, then $\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2)$ (corollary 10.6).

Consider the vector space $V := W_1 \times W_2$:[10] This is the set of pairs $(\vec{w}_1, \vec{w}_2)$, with $\vec{w}_1 \in W_1$ and $\vec{w}_2 \in W_2$. Addition and scalar multiplication are defined entry by entry (just like for K^n). Note that $\dim(V) = \dim(W_1) + \dim(W_2)$, because if $(\vec{v}_1, \ldots, \vec{v}_r)$ is a basis of W_1 and $(\vec{w}_1, \ldots, \vec{w}_s)$ is a basis of W_2, then the family $((\vec{v}_i, \vec{w}_j))_{(i,j)\in\{1,\ldots,r\}\times\{1,\ldots,s\}}$ is a basis of V.

Let $f : V \to W$ be the map sending $(\vec{w}_1, \vec{w}_2)$ to $\vec{w}_1 - \vec{w}_2$. I claim that $\mathrm{Im}(f) = W_1 + W_2$ and $\mathrm{Ker}(f) = W_1 \cap W_2$ (we see this as a subspace of V by sending $\vec{w} \in W_1 \cap W_2$ to $(\vec{w}, \vec{w}) \in V$). This immediately implies the dimension formula by the rank-nullity theorem. Let's first calculate $\mathrm{Ker}(f)$. Clearly, if $\vec{w} \in W_1 \cap W_2$, then $f(\vec{w}, \vec{w}) = \vec{w} - \vec{w} = 0$. Let $(\vec{w}_1, \vec{w}_2) \in \mathrm{Ker}(f)$. Then $\vec{w}_1 - \vec{w}_2 = 0$, so $\vec{w}_1 = \vec{w}_2 \in W_1 \cap W_2$. Now let's calculate $\mathrm{Im}(f)$. If $\vec{v} \in W_1 + W_2$, write $\vec{v} = \vec{w}_1 + \vec{w}_2$ with $\vec{w}_1 \in W_1$ and $\vec{w}_2 \in W_2$; then $\vec{v} = f(\vec{w}_1, -\vec{w}_2) \in \mathrm{Im}(f)$. If $\vec{v} \in \mathrm{Im}(f)$, write $\vec{v} = f(\vec{w}_1, \vec{w}_2)$; then $\vec{v} = \vec{w}_1 + (-\vec{w}_2) \in W_1 + W_2$.

Remark 16.4 The corollary to the rank-nullity theorem is totally false for infinite-dimensional vector spaces. Here are counterexamples.

[10] This vector space is also sometimes denoted $W_1 \oplus W_2$, but I won't do this here to avoid confusion. Note however that the notation is coherent : With this notation, saying that W_1 and W_2 are in direct sum is equivalent to saying that the map $W_1 \oplus W_2 \to W_1 + W_2$, $(\vec{w}_1, \vec{w}_2) \mapsto \vec{w}_1 + \vec{w}_2$ is an isomorphism. (Please ignore this remark if you find it confusing.)

Let's take $V = W = K[X]$. First consider $f : V \to V,\ P \mapsto XP$. Then f is linear and injective (because $\mathrm{Ker}(f) = \{0\}$), but it's not surjective; in fact, $\mathrm{Im}(f)$ is the subspace of polynomials with zero constant term. Now let's consider $g : V \to V,\ P \mapsto P'$. Then g is linear and surjective (every polynomials can be written as the derivative of another polynomial), but not injective; indeed, the constant polynomials are in $\mathrm{Ker}(f)$, so $\mathrm{Ker}(f) = \{0\}$.

3/7/2017

17 Dual space

At the beginning of the semester, I said that we would give a linear algebraic reinterpretation of the transpose of a matrix. Now the time has come.

We fix a field K and a K-vector space V.

Definition 17.1 (1) A *linear form* (or *linear functional*) on V is a linear transformation from V to K.

(2) The *dual (space)* of V is the K-vector space $V^* = \mathrm{Hom}(V, K)$. In other words, it's the space of linear forms on V.

(3) $V^{**} := (V^*)^*$ is called the *bidual* of V.

Remark 17.2 By remark 14.4, if V is finite-dimensional, then :

$$\dim(V^*) = (\dim V)(\dim K) = \dim V.$$

If $f \in V^*$ and $\vec{v} \in V$, we sometimes write $\langle f, \vec{v} \rangle$ instead of $f(\vec{v})$.

Lemma 17.3 *The map $\iota : V \to V^{**}$ that sends $\vec{v} \in V$ to the linear form on $\iota : V^* \to K$ given by $\iota(f) = f(\vec{v})$ is a linear transformation. If V is finite-dimensional, this linear transformation is injective, hence it is an isomorphism.*

Proof. We actually have three things to prove here :

(A) $\iota(\vec{v}) : V^* \to K$ is a linear map for every $\vec{v} \in V$.

(B) The map $\iota : V \to V^{**}$ is linear.

(C) If V is finite-dimensional, then the map ι is injective.

Note that the last part (the fact that ι is an isomorphism if V is finite-dimensional) follows from the fact that ι is injective, the fact that $\dim(V^{**}) \dim(V^*) = \dim(V)$ (see remark 17.2), and corollary 16.2.

Let's prove (A). Let $\vec{v} \in V$. If $f_1, f_2 \in V^*$ and $\lambda \in K$, then

$$(\iota(\vec{v}))(f_1 + f_2) = (f_1 + f_2)(\vec{v}) = f_1(\vec{v}) + f_2(\vec{v}) = (\iota(\vec{v}))(f_1) + (\iota(\vec{v}))(f_2)$$

and

$$(\iota(\vec{v}))(\lambda f) = (\lambda f)(\vec{v}) = \lambda f(\vec{v}) = \lambda(\iota(\vec{v}))(f).$$

[11]

Let's prove (B). Let $\vec{v}_1, \vec{v}_2 \in V$ and $\lambda \in K$. Then, for every $f \in V^*$,

$$(\iota(\vec{v}_1 + \vec{v}_2))(f) = f(\vec{v}_1 + \vec{v}_2) = f(\vec{v}_1) + f(\vec{v}_2) = (\iota(\vec{v}_1))(f) + (\iota(\vec{v}_2))(f)$$

and

$$(\iota(\lambda\vec{v}))(f) = f(\lambda\vec{v}) = \lambda f(\vec{v}) = \lambda(\iota(\vec{v}))(f),$$

so $\iota(\vec{v}_1 + \vec{v}_2) = \iota(\vec{v}_1) + \iota(\vec{v}_2)$ and $\iota(\lambda\vec{v}) = \lambda\iota(\vec{v})$.

Let's prove (C). By proposition 15.4, it's enough to prove that $\mathrm{Ker}(\iota) = \{0\}$. Suppose that $\mathrm{Ker}(\iota) \neq \{0\}$, and choose $\vec{v} \in \mathrm{Ker}(\iota)$ such that $\vec{v} \neq 0$. Then (by (iii) of theorem 9.3) we can find $\vec{v}_2, \ldots, \vec{v}_n \in V$ such that $(\vec{v}, \vec{v}_2, \ldots, \vec{v}_n)$ is a basis of V. By (ii) of theorem 14.1, there exists a unique linear transformation $f : V \to K$ such that $f(\vec{v}) = 1$ and $f(\vec{v}_2) = \cdots = f(\vec{v}_n) = 0$. But then $(\iota(\vec{v}))(f) = f(\vec{v}) = 1$, which contradicts the fact that $\vec{v} \in \mathrm{Ker}(\iota)$. So we must have $\mathrm{Ker}(\iota) = \{0\}$.

□

Remark 17.4 What happens if V is infinite-dimensional ? If we admit the axiom of choice, we are also able to complete any nonzero vector of V to a basis, and then the proof of (C) works and shows that ι is injective. However, if we don't admit the axiom the choice, then the existence of vector spaces V such that ι is not injective does not contradict the other axioms of set theory (the Zermelo-Fraenkel axioms).

Example 17.5 If $V = K^3$, let's calculate V^*. One element of V^* is the linear transformation $T_1 : V \to K$, $(x_1, x_2, x_3) \mapsto x_1$. Others are $T_2 : V \to K$, $(x_1, x_2, x_3) \mapsto x_2$ and $T_3 : V \to K$, $(x_1, x_2, x_3) \mapsto x_3$.

In fact, these three form a basis of V^*. Why ? Well, if we think of V as the space of column matrices with 3 rows, then we have seen in example 14.8 that V^*, which is $\mathrm{Hom}(V, K)$, can be identified with $M_{1,3}(K)$, and then the map $V^* \times V \to K$, $(f, \vec{v}) \mapsto f(\vec{v})$, is just matrix multiplication. In that identification, T_1 corresponds to the matrix $\begin{pmatrix}1 & 0 & 0\end{pmatrix}$, T_2 corresponds to the matrix $\begin{pmatrix}0 & 1 & 0\end{pmatrix}$ and T_3 corresponds to the matrix $\begin{pmatrix}0 & 0 & 1\end{pmatrix}$. It is clear that these three matrices form a basis of $M_{1,3}(K)$.

Here is another way to show that (T_1, T_2, T_3) is a basis of V^*. Let $(\vec{e}_1, \vec{e}_2, \vec{e}_3)$ be the canonical basis of V. First we show that T_1, T_2 and T_3 are linearly independent : Let $\lambda_1, \lambda_2, \lambda_3 \in K$ such that the linear transformation $T = \lambda_1 T_1 + \lambda_2 T_2 + \lambda_3 T_3$ is equal to 0. Then $0 = T(\vec{e}_1) = \lambda_1$, and similarly $\lambda_2 = T(\vec{e}_2) = 0$ and $\lambda_3 = T(\vec{e}_3) = 0$. Now let's show that $V^* = \mathrm{Span}(T_1, T_2, T_3)$. Let $T \in V^*$, and let $U = T - (T(\vec{e}_1)T_1 + T(\vec{e}_2)T_2 + T(\vec{e}_3)T_3)$. Then $U(\vec{e}_1) = U(\vec{e}_2) = U(\vec{e}_3) = 0$, so $U = 0$, so $T = T(\vec{e}_1)T_1 + T(\vec{e}_2)T_2 + T(\vec{e}_3)T_3 \in \mathrm{Span}(T_1, T_2, T_3)$.

[11] Note that we have not used the fact that the maps f_1 and f_2 are linear. In fact, if $\vec{v} \in V$, the map $K^V \to K$, $f \mapsto f(\vec{v})$, is linear; $\iota(\vec{v})$ is just the restriction of this map to $\mathrm{Hom}(V, K) \subset K^V$.

Example 17.6 Let's consider the subspace V of $\mathbb{Q}^3$ defined by the equation $x+y+z=0$. What does the dual of V look like ?

First note that the element $T_1, T_2, T_3 \in (\mathbb{Q}^3)^*$ of the previous example can be retricted to V and give elements $U_1, U_2, U_3 \in V^*$. It seems logical that these should generate V^*, but the proof of the previous example doesn't work because neither $\vec{e}_1$ nor $\vec{e}_2$ nor $\vec{e}_3$ are in V. So let's try something else. By theorem 10.5, there exists a subspace W of $\mathbb{Q}^3$ such that $\mathbb{Q}^3 = V \oplus W$. (In fact it is easy to construct such a W, we can take for example W the line generated by the vector $(1,1,1)$.) Let $U \in V^*$. We define $T : \mathbb{Q}^3 \to \mathbb{Q}$ in the following way : If $\vec{v} \in \mathbb{Q}^3$, write $\vec{v} = \vec{v}_1 + \vec{v}_2$ with $\vec{v}_1 \in V$ and $\vec{v}_2 \in W$ and set $T(\vec{v}) = U(\vec{v}_1)$; this makes sense because the decomposition of $\vec{v}$ is unique, and it's linear for the same reason.

For example, taking $W = \mathrm{Span}((1,1,1))$, let's calculate what this gives for $U = U_2$. If $\vec{v} = (x,y,z) \in \mathbb{Q}^3$, then $\vec{v} = \vec{v}_1 + \vec{v}_2$, with $\vec{v}_1 = (x-(x+y+z)/3, y-(x+y+z)/3, z-(x+y+z)/3) \in V$ and $\vec{v}_2 = ((x+y+z)/3, (x+y+z)/3, (x+y+z)/3) \in W$. Then the corresponding T sends (x,y,z) to $y-(x+y+z)/3 = -x/3 + 2y/3 - z/3$.

Anyway, if $U \in V^*$, we have shown how to find $T \in (\mathbb{Q}^3)^*$ such that $U = T_{|V}$. By the previous example, we can write $T = a_1T_1 + a_2T_2 + a_3T_3$ with $a_1, a_2, a_3 \in \mathbb{Q}$, and then $U = a_1U_1 + a_2U_2 + a_3U_3$.

We next have to see if U_1, U_2, U_3 are linearly independent, but this cannot be true because $U_3 = -U_1 - U_2$. So let's see if U_1 and U_2 are linearly independent. Let $a_1, a_2 \in \mathbb{Q}$ such that $U := a_1U_1 + a_2U_2 = 0$. Then $U((1,-1,0)) = a_1 - a_2 = 0$ and $U((1,0,-1)) = a_1 = 0$, so $a_1 = 0$ and $a_2 = 0$. So finally, we get that (U_1, U_2) is a basis of V^*.

Example 17.7 Let I be a set. Then the dual of $K^{(I)}$ is isomorphic to K^I in a natural way. Indeed, consider the map $\varphi : K^I \to (K^{(I)})^*$ that sends $f : I \to K$ to the function $\varphi(f) : K^{(I)} \to K,\ g \mapsto \sum_{i\in I} f(i)g(i)$ (the sum is finite because g has finite support, so it makes sense). This map u is linear (exercise). Before we continue, remember that we have the canonical basis $(e_i)_{i\in I}$ of $K^{(I)}$ defined in example 8.3.

Let's show that φ is injective, which amounts to showing that $\mathrm{Ker}(\varphi) = \{0\}$. Let $f \in \mathrm{Ker}(\varphi)$. Then, for every $i \in I$,

$$(\varphi(f))(e_i) = \sum_{j\in I} f(j)e_i(j) = f(i) = 0.$$

So $f = 0$.

Let's show that φ is surjective. Let $u \in (K^{(I)})^*$. We define $f : I \to K$ by $f(i) = u(e_i)$. Then, for every $i \in I$, $(\varphi(f))(e_i) = f(i) = u(e_i)$. As $(e_i)_{i\in I}$ is a basis of $K^{(I)}$, this shows (by theorem 14.1) that $\varphi(f) = u$.

In particular, we get an isomorphism $(K^n)^* \xrightarrow{\sim} K^n$. Note that the map $(K^n)^* \times K^n \to K,\ (f, \vec{v}) \mapsto f(\vec{v})$ then becomes $K^n \times K^n \to K,\ ((x_1,\dots,x_n),(y_1,\dots,y_n)) \mapsto x_1y_1 + \cdots + x_ny_n$.

Definition 17.8 (Dual basis.) Suppose that V is finite-dimensional, and let $(\vec{v}_1,\dots,\vec{v}_n)$ be a basis of V. Define a family of vectors $(\vec{v}_1^*,\dots,\vec{v}_n^*)$ of V^* in the following way : For

every $i, j \in \{1, \ldots, n\}$,

$$\vec{v}_i^*(\vec{v}_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

(This uniquely determines the $\vec{v}_i^*$ by theorem 14.1.)

This family is called the *dual basis* of the basis $(\vec{v}_1, \ldots, \vec{v}_n)$.

Remark 17.9 We are still in the situation of definition 17.8. Write $\mathfrak{B} = (\vec{v}_1, \ldots, \vec{v}_n)$. If $\vec{v} \in V$, we can write $\vec{v} = \lambda_1\vec{v}_1 + \cdots + \lambda_n\vec{v}_n$. Then, for every $i \in \{1, \ldots, n\}$,

$$\vec{v}_i(\vec{v}) = \vec{v}_i^*(\lambda_1\vec{v}_1 + \cdots + \lambda_n\vec{v}_n) = \vec{v}_i^*(\lambda_i\vec{v}_i) = \vec{v}_i.$$

In other words, we get the following formula :

$$[\vec{v}_{\mathfrak{B}}] = (\vec{v}_1^*(\vec{v}), \ldots, \vec{v}_n^*(\vec{v})).$$

The name "dual basis" is justified by the following result :

Proposition 17.10 *With the notation and assumptions of definition 17.8, $(\vec{v}_1^*, \ldots, \vec{v}_n^*)$ is a basis of V^*.*

Note that this reproves the fact that V^* is finite-dimensional and $\dim(V^*) = \dim(V)$.

Proof. Let's show that this family is linearly independent. Let $\lambda_1, \ldots, \lambda_n \in K$ be such that $\lambda_1\vec{v}_1^* + \cdots + \lambda_n\vec{v}_n^* = 0$. Then, for every $i \in \{1, \ldots, n\}$,

$$\lambda_i = \lambda_i\vec{v}_i^*(\vec{v}_i) = (\lambda_1\vec{v}_1^* + \cdots + \lambda_n\vec{v}_n^*)(\vec{v}_1) = 0,$$

Let's show that this family generates V^*. Let $f \in V^*$, and let $\lambda_i = f(\vec{v}_i)$ for $i \in \{1, \ldots, n\}$. Then, for every i,

$$(f - (\lambda_1\vec{v}_1^* + \cdots + \lambda_n\vec{v}_n^*))(\vec{v}_i) = f(\vec{v}_i) - \lambda_i\vec{v}_i^*(\vec{v}_i) = 0.$$

By theorem 14.1, $f = \lambda_1\vec{v}_1^* + \cdots + \lambda_n\vec{v}_n^*$.

□

Example 17.11 The basis (T_1, T_2, T_3) of 17.5 is the dual basis of the canonical basis.

Example 17.12 In example 17.6, we found a basis (U_1, U_2) of V^*, where V is the subspace of $\mathbb{Q}^3$ defined by the equation $x + y + z = 0$. We have a basis $(\vec{v}_1, \vec{v}_2)$ of V given by $\vec{v}_1 = (1, -1, 0)$ and $\vec{v}_2 = (1, 0, -1)$. Note that (U_1, U_2) is not the dual basis of $(\vec{v}_1, \vec{v}_2)$, because $U_1(\vec{v}_1) = 1$ but $U_2(\vec{v}_1) = -1$. So what is the dual basis of $(\vec{v}_1, \vec{v}_2)$?

Well, we need two linear transformations $\vec{v}_1^*, \vec{v}_2^* : V \to K$ such that $\vec{v}_1^*(\vec{v}_1) = \vec{v}_2^*(\vec{v}_2) = 1$ and $\vec{v}_1^*(\vec{v}_2) = \vec{v}_2^*(\vec{v}_1) = 0$. If we write $\vec{v}_1^*$ in the form $(x, y, z) \mapsto ax + by + cz$, this means that we must have $a - b = 1$ and $b - c = 0$, hence $b = c$ and $a = 1 + b$. So we can take $b = c = 0$ and $a = 1$, which gives $\vec{v}_1^* = U_1 : (x, y, z) \mapsto x$. For $\vec{v}_2^*$, if we write it as $(x, y, z) \mapsto ax + by + cz$, then we must have $a - c = 1$ and $a - b = 0$, hence $a = b$ and $a = 1 + c$. So we can take $c = 0$ and $a = b = 1$, which gives $\vec{v}_2^* = U_1 + U_2 : (x, y, z) \mapsto x + y$.

Example 17.13 If we identity $(K^n)^*$ to K^n as in example 17.7 (see also example 17.5 for $n = 3$), then the dual basis of the canonical basis is just the canonical basis.

Corollary 17.14 *If V is finite-dimensional, then the map $\iota : V \to V^{**}$ is an isomorphism. Moreover, for every basis $(\vec{v}_1, \ldots, \vec{v}_n)$, the basis $(\iota(\vec{v}_1), \ldots, \iota(\vec{v}_n))$ of V^{**} is the dual basis of $(\vec{v}_1^*, \ldots, \vec{v}_n^*)$.*

Proof. We know that ι is injective, and we also know that $\dim(V^{**}) = \dim(V^*) = \dim(V)$. By corollary 16.2, this implies that ι is an isomorphism.

Let $(\vec{v}_1, \ldots, \vec{v}_n)$ be a basis of V. Then, for every $i, j \in \{1, \ldots, n\}$,

$$(\iota(\vec{v}_i))(\vec{v}_j^*) = \vec{v}_j^*(\vec{v}_i) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

This proves the second sentence.

□

Definition 17.15 (Transpose of a linear transformation.) Let V and W be two vector spaces, and let $u : V \to W$ be a linear transformation.

Then the map ${}^t u : W^* \to V^*$ sending $f \in W^*$ to $f \circ u$ is a linear transformation, called the *transpose* of u.

That ${}^t u$ is linear follows directly from the properties of composition (see lemma 13.4).

Example 17.16 Consider the linear transformation $u : K^3 \to K^2$ given by the matrix $A = \begin{pmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \end{pmatrix}$. What is ${}^t u$?

We identify K^3 and K^2 to their duals as in example 17.13, so we can also look at the matrix of ${}^t u$ in the canonical bases of these spaces.

Let $\vec{v} = (a, b) \in K^2$. As an element of $(K^2)^*$, this is the linear transformation $a\vec{e}_1^* + b\vec{e}_2^*$, which sends (x, y) to $ax + by$. So ${}^t u(\vec{v}) \in (K^3)^*$ is the linear transformation :

$$(x, y, z) \mapsto u(x, y, z) = (y+2z, 3x+4y+5z) \mapsto a(y+2z)+b(3x+4y+5z) = (3b)x+(a+4b)y+(2a+5b)z.$$

In other words, ${}^t u(\vec{v}) = (3b)\vec{e}_1^* + (a + 4b)\vec{e}_2^* + (2a + 5b)\vec{e}_3^*$. This means that the matrix of ${}^t u$ in the duals bases of the canonical bases is $\begin{pmatrix} 0 & 3 \\ 1 & 4 \\ 2 & 5 \end{pmatrix}$, which is A^T. This is not a coincidence.

Proposition 17.17 *Let V and W be finite-dimensional vector spaces. Fix a basis $\mathfrak{A}$ of V and a basis $\mathfrak{B}$ of W, and denote by $\mathfrak{A}^*$ and $\mathfrak{B}^*$ the dual bases.*

Then for every linear transformation $u : V \to W$, we have

$$[{}^t u]_{\mathfrak{A}^*, \mathfrak{B}^*} = [u]_{\mathfrak{B}, \mathfrak{A}}^T.$$

Proof. Let $A = [u]_{\mathfrak{B},\mathfrak{A}}$ and $B = [{}^t u]_{\mathfrak{A}^*,\mathfrak{B}^*}$. Write $\mathfrak{A} = (\vec{v}_1, \dots, \vec{v}_n)$ and $\mathfrak{B} = (\vec{w}_1, \dots, \vec{w}_m)$. Remember that, by definition, A is the matrix with columns $[u(\vec{v}_1)]_{\mathfrak{B}}, \dots, [u(\vec{v}_n)]_{\mathfrak{B}}$. By remark 17.9, the $(j,i)th$ of A is

$$\vec{w}_j^*(u(\vec{v}_i)) = (\vec{w}_j^* \circ u)(\vec{v}_i) = (\iota(\vec{v}_i))({}^t u(\vec{w}_j^*)).$$

By corollary 17.14 (and remark 17.9 again, this times used to calculate the entries of B), this is exactly the (i,j)th entry of B. So we indeed have $B = A^T$.

□

3/9/2017

18 Determinants 1 : multilinear forms

You might have met the notion of the determinant of a matrix before. It's a number, calculated in some mysterious and complicated way, that tells you whether a matrix is invertible. And for a 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the determinant is simply $ad - bc$.

The goal of this section (and the following ones) is to explain the theory behind determinants.

Let K be a field and V be a K-vector space.

Definition 18.1 A *bilinear form* on V is a map $f : V \times V \to K$ that is linear in each variable, that is :

- For every $\vec{v}_1, \vec{v}_2, \vec{w} \in V$, $f(\vec{v}_1 + \vec{v}_2, \vec{w}) = f(\vec{v}_1, \vec{w}) + f(\vec{v}_2, \vec{w})$.
- For every $\vec{v}, \vec{w}_1, \vec{w}_2 \in V$, $f(\vec{v}, \vec{w}_1 + \vec{w}_2) = f(\vec{v}, \vec{w}_1) + f(\vec{v}, \vec{w}_2)$.
- For every $\vec{v}, \vec{w} \in V$, for every $\lambda \in K$, $f(\lambda\vec{v}, \vec{w}) = f(\vec{v}, \lambda\vec{w}) = \lambda f(\vec{v}', \vec{w})$.

More generally, if $n \geq 2$, a *n-linear form* on V is a map $f : V^n \to K$ that is linear in each variable. That is, for every $i \in \{1, \dots, n\}$ and for every $\vec{v}_1, \dots, \vec{v}_{i-1}, \vec{v}_{i+1}, \dots, \vec{v}_n \in V$, the map $\vec{v} \mapsto f(\vec{v}_1, \dots, \vec{v}_{i-1}, \vec{v}, \vec{v}_{i+1}, \dots, \vec{v}_n)$ is a linear form on V.

Remark 18.2 If W is another K-vector space, we can define similarly n-linear maps from V^n to W. Even more generally, if $V_1, \dots, V_n, W$ are K-vector spaces, we can talk about n-linear (or multilinear) maps from $V_1 \times \cdots \times V_n$ to W.

Example 18.3 (1) If $V = K^n$, the following map is a bilinear form on V :

$$f : ((x_1, \dots, x_n), (y_1, \dots, y_n)) \mapsto x_1 y_1 + \cdots + x_n y_n.$$

Note that we have $f(\vec{v}, \vec{w}) = f(\vec{w}, \vec{v})$. This property of f is called being *symmetric*.

(2) If $V = K^2$, the map $f : ((x_1, x_2), (y_1, y_2)) \mapsto x_1 y_2 - x_2 y_1$ is a bilinear form on V. Note that we have $f(\vec{v}, \vec{v}) = 0$. (Such a f is called *alternating*.)

(3) If $V = K[X]$ and $x_1, \dots, x_n \in K$, the map $k : (P_1, \dots, P_n) \mapsto P_1(x_1) \dots P_n(x_n)$ is a n-linear form on V.

(4) Suppose that $f_1, \dots, f_n \in V^*$. Then the map $f : V^n \to K$ sending $(\vec{v}_1, \dots, \vec{v}_n)$ to $f_1(\vec{v}_1) \dots f_n(\vec{v}_n)$ is a n-linear form. We'll denote it by $f_1 \otimes \dots \otimes f_n$.

We'll generalize the property of example (2).

Definition 18.4 A n-linear form f on V is called *alternating* if $f(\vec{v}_1, \dots, \vec{v}_n) = 0$ as soon as there exist $i \neq j$ such that $\vec{v}_i = \vec{v}_j$.

Note that alteranting n-linear forms form a subspace of the space of n-linear forms (that is, a linear combination of alternating n-linear forms stays alternating).

Here are some more basic properties of alternating forms :

Lemma 18.5 *Let f be an alternating n-linear form on V.*

(i) For all $\vec{v}_1, \dots, \vec{v}_n \in V$ and $1 \leq i < j \leq n$, we have

$$f(\vec{v}_1, \dots, \vec{v}_{i-1}, \vec{v}_j, \vec{v}_{i+1}, \dots \vec{v}_{j-1}, \vec{v}_i, \vec{v}_{j+1}, \dots, \vec{v}_n) = -f(\vec{v}_1, \dots, \vec{v}_n).$$

(ii) Let $\vec{v}_1, \dots, \vec{v}_n \in V$, pick $i \in \{, \dots, n\}$ and $\lambda_j \in K$, $j \neq i$. Then :

$$f(\vec{v}_1, \dots, \vec{v}_{i-1}, \vec{v}_i + \sum_{j \neq i} \lambda_j \vec{v}_j, \vec{v}_{i+1}, \dots, \vec{v}_n) = f(\vec{v}_1, \dots, \vec{v}_n).$$

(iii) If $(\vec{v}_1, \dots, \vec{v}_n)$ is not linearly independent, then $f(\vec{v}_1, \dots, \vec{v}_n) = 0$.

Proof.

(i) Let's consider the $\vec{v}_r, r \neq i, j$ as fixed and write $g(\vec{v}_i, \vec{v}_j)$ for $f(\vec{v}_1, \dots, \vec{v}_n)$.

So we know that $g(\vec{v}_i, \vec{v}_j) = 0$ if $\vec{v}_i = \vec{v}_j$, and we're trying to prove that $g(\vec{v}_j, \vec{v}_i) = -g(\vec{v}_i, \vec{v}_j)$. For this we calculate, using the linearity in each variable :

$$0 = g(\vec{v}_i + \vec{v}_j, \vec{v}_i + \vec{v}_j) = g(\vec{v}_i, \vec{v}_i) + g(\vec{v}_i, \vec{v}_j) + g(\vec{v}_j, \vec{v}_i) + g(\vec{v}_j, \vec{v}_j) = g(\vec{v}_i, \vec{v}_j) + g(\vec{v}_j, \vec{v}_i),$$

which gives the conclusion.

(ii) We have $f(\vec{v}_1, \dots, \vec{v}_{i-1}, \vec{v}_i + \sum_{j \neq i} \lambda_j \vec{v}_j, \vec{v}_{i+1}, \dots, \vec{v}_n) =$

$$f(\vec{v}_1, \dots, \vec{v}_n) + \sum_{j \neq i} \lambda_j f(\vec{v}_1, \dots, \vec{v}_{i-1}, \vec{v}_j, \vec{v}_{i+1}, \dots, \vec{v}_n).$$

All the terms in the sum except the first one are zero, because f is alternating.

(iii) If $(\vec{v}_1, \ldots, \vec{v}_n)$ is not linearly independent, then we have $\lambda_1, \ldots, \lambda_n \in K$ not all zero such that $\lambda_1\vec{v}_1 + \cdots + \lambda_n\vec{v}_n = 0$. Fix $i \in \{1, \ldots, n\}$ such that $\lambda_i \neq 0$, and write $\mu_j = -\lambda_i^{-1}\lambda_j$ for $j \neq i$. Then $\vec{v}_i = \sum_{1\leq j\leq n, j\neq i} \mu_j\vec{v}_j$. Hence :

$$\begin{aligned} f(\vec{v}_1, \ldots, \vec{v}_n) &= f(\vec{v}_1, \ldots, \vec{v}_{i-1}, \sum_{j\neq i} \mu_j\vec{v}_j, \vec{v}_{i+1}, \ldots, \vec{v}_n) \\ &= \sum_{j\neq i} \mu_j f(\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_j, \vec{v}_{i+1}, \ldots, \vec{v}_n) \\ &= 0 \end{aligned}$$

(By definition of "alternating", every term in the last sum is zero.)

□

Remark 18.6 If $\mathrm{char}(K) \neq 2$ (which means that 2 is invertible in K), then any n-linear form that satisfies part (i) of the lemma is alternating.

To generalize point (i) of the lemma, we need a more efficient way to talk about permutations of the indices $1, \ldots, n$.

19 Determinants 2 : the symmetric group $\mathfrak{S}_n$

Definition 19.1 The *symmetric group* $\mathfrak{S}_n$ is the set of bijections $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$. If $\sigma, \tau \in \mathfrak{S}_n$, we usually write $\sigma\tau$ instead of $\sigma \circ \tau$. We also write 1 for the identity of $\{1, \ldots, n\}$.

Remark 19.2 $\mathfrak{S}_n$ is a group in the sense of definition 1.4. Remember that this just means that it has a multiplication (here it's composition), that the multiplication is associative (which is true for composition), that there is an identity element (here it's given by the identity map of $\{1, \ldots, n\}$), and that each element has an inverse for the multiplication (which is true here because all the elements of $\mathfrak{S}_n$ are assumed to be bijections).

Remark 19.3

- We have

$$|\mathfrak{S}_n| = n! := n \times (n-1) \times \cdots \times 2 \times 1.$$

- There is an often-used short notation for elements of $\mathfrak{S}_n$. It's best explained on examples.

  For example, $(354) \in \mathfrak{S}_6$ is the element σ given by $\sigma(1) = 1$, $\sigma(2) = 2$, $\sigma(3) = 5$, $\sigma(4) = 3$, $\sigma(5) = 4$, $\sigma(6) = 6$.

  Or $(14)(23) \in \mathfrak{S}_4$ is the element σ given by $\sigma(1) = 4$, $\sigma(2) = 3$, $\sigma(3) = 2$, $\sigma(4) = 1$.

  Or $(12\ldots n) \in \mathfrak{S}_n$ is the element σ given by $\sigma(i) = i+1$ if $i \leq n-1$ and $\sigma(n) = 1$.

Remark 19.4 We see $\mathfrak{S}_n$ as a subset[12] of $\mathfrak{S}_{n+1}$ in the following way : If $\sigma \in \mathfrak{S}_n$, we can see it as the element of $\mathfrak{S}_{n+1}$ that sends each $i \leq n$ to $\sigma(i)$ and sends $n+1$ to $n+1$.

Definition 19.5 An element of $\mathfrak{S}_n$ of the form $\sigma = (ij)$ for $i \neq j$[13] is called a *transposition.*

Note that if σ is a transposition, then $\sigma^{-1} = \sigma$.

Proposition 19.6 *Let $n \geq 1$. Then any $\sigma \in \mathfrak{S}_n$ is a product of transpositions. (In general, in many different ways.)*

By convention, we think of 1 as the product of zero transpositions.

Proof. We do an induction on n. If $n = 1$, then $\mathfrak{S}_n = \{1\}$ and the result is obvious. Suppose that $n \geq 2$ and that the result is known for $n-1$. Let $\sigma \in \mathfrak{S}_n$. If $\sigma(n) = n$, then we can see σ as an element of $\mathfrak{S}_{n-1}$ as in remark 19.4 and use the induction hypothesis. Otherwise, let $i = \sigma(n)$, and let $\tau = (in)\sigma$. Then $\tau(n) = n$, so, as before, τ is a product of transpositions by the induction hypothesis. Then as $\sigma = (in)\tau$ (remember that $(in)^{-1} = (in)$), σ is also a product of transpositions.

□

The second thing we will need about the symmetric group is the definition of the sign of a permutation.

Definition 19.7 Let $\sigma \in \mathfrak{S}_n$. Then its *sign* (or *signature*) is defined by :

$$\mathrm{sgn}(\sigma) = (-1)^{|\{(i,j)\in\{1,\ldots,n\}^2 | i<j \text{ and } \sigma(j)>\sigma(i)\}|}.$$

Example 19.8 If $\sigma = 1$, then $\mathrm{sgn}(\sigma) = 1$.

Example 19.9 If σ is a transposition, then $\mathrm{sgn}(\sigma) = -1$.

Indeed, suppose that $\sigma = (ij)$ with $i < j$. We have to find all the pairs $(a, b) \in \{1, \ldots, n\}^2$ such that $a < b$ and $\sigma(a) > \sigma(b)$. This is only possible if at least one of a, b is in $\{i, j\}$, and we get three cases :

(a) $(a, b) = (i, j)$.

(b) $a = i$ and $i < b < j$ ($j - i - 1$ possible pairs (a, b)).

(c) $i < a < j$ and $b = j$ ($j - i - 1$ possible pairs (a, b)).

So $\mathrm{sgn}(\sigma) = (-1)^{1+2(j-i-1)} = -1$.

For theoretical purposes, the following formula is more convenient :

[12] Actually a subgroup, which means a subset containing 1 and stable by multiplication and inversion.
[13] This means that $\sigma(i) = j$, $\sigma(j) = i$, and $\sigma(k) = k$ if $k \neq i, j$.

Proposition 19.10 *Let $\sigma \in \mathfrak{S}_n$. Then*

$$\operatorname{sgn}(\sigma) = \prod_{1 \leq i < j \leq n} \frac{\sigma(i) - \sigma(j)}{i - j}.$$

Proof. Let $\sigma \in \mathfrak{S}_n$, and write sgn' for $\prod_{1 \leq i < j \leq n} \frac{\sigma(i) - \sigma(j)}{i - j}$. Note that $\operatorname{sgn}' \in \mathbb{Q}$.

We also have

$$(\operatorname{sgn}')^2 = \left(\prod_{i<j} \frac{\sigma(i) - \sigma(j)}{i - j} \right)^2 = \frac{(\prod_{i<j}(\sigma(i) - \sigma(j)))^2}{(\prod_{i<j}(i - j))^2}.$$

If $i < j$, then $(i - j)^2 = (j - i)^2$ and $(\sigma(i) - \sigma(j))^2 = (\sigma(j) - \sigma(i))^2$. So we also have :

$$(\operatorname{sgn}')^2 = \frac{(\prod_{i>j}(\sigma(i) - \sigma(j)))^2}{(\prod_{i>j}(i - j))^2}.$$

This implies that :

$$(\operatorname{sgn}')^4 = \frac{\prod_{i \neq j}(\sigma(i) - \sigma(j))^4}{\prod_{i \neq j}(i - j)^4}.$$

As σ is a bijection, $(\operatorname{sgn}')^4 = 1$. As $\operatorname{sgn}' \in \mathbb{Q}$, $\operatorname{sgn}' \in \{\pm 1\}$.

So we only need to show that sgn' has the same sign as $\operatorname{sgn}(\sigma)$. But the sign of sgn' is clearly (-1) raised to the number of couples (i, j) such that $i < j$ and $\sigma(i) > \sigma(j)$, so we are done.

□

Corollary 19.11 *For every $\sigma, \tau \in \mathfrak{S}_n$, $\operatorname{sgn}(\sigma\tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)$.* [14]

Proof. We have :

$$\begin{aligned}
\operatorname{sgn}(\sigma\tau) &= \prod_{1 \leq i < j \leq n} \frac{\sigma\tau(i) - \sigma\tau(j)}{i - j} \\
&= \prod_{1 \leq i < j \leq n} \frac{\sigma\tau(i) - \sigma\tau(j)}{\tau(i) - \tau(j)} \frac{\tau(i) - \tau(j)}{i - j} \\
&= \prod_{1 \leq i < j \leq n} \frac{\sigma\tau(i) - \sigma\tau(j)}{\tau(i) - \tau(j)} \prod_{1 \leq i \leq n} \frac{\tau(i) - \tau(j)}{i - j} \\
&= \operatorname{sgn}(\tau) \prod_{1 \leq i < j \leq n} \frac{\sigma(\tau(i)) - \sigma(\tau(j))}{\tau(i) - \tau(j)}
\end{aligned}$$

To finish the proof, we need to show that :

$$\operatorname{sgn}(\sigma) = \prod_{1 \leq i < j \leq n} \frac{\sigma(\tau(i)) - \sigma(\tau(j))}{\tau(i) - \tau(j)}.$$

[14] We say that sgn is a *morphism of groups* from $\mathfrak{S}_n$ to $\{\pm 1\}$.

Let $\alpha, \beta \in \{1, \ldots, n\}$ such that $\alpha < \beta$. There are unique numbers $i, j \in \{1, \ldots, n\}$ such that $i < j$ and $\{\tau(i), \tau(j)\} = \{\alpha, \beta\}$. (Note that this is an equality of sets, not of ordered pairs.) There are two possibilities :

- If $\tau(i) = \alpha$ and $\tau(j) = \beta$, then :

$$\frac{\sigma(\tau(i)) - \sigma(\tau(j))}{\tau(i) - \tau(j)} = \frac{\sigma(\alpha) - \sigma(\beta)}{\alpha - \beta}.$$

- If $\tau(i) = \beta$ and $\tau(j) = \alpha$, them :

$$\frac{\sigma(\tau(i)) - \sigma(\tau(j))}{\tau(i) - \tau(j)} = \frac{\sigma(\beta) - \sigma(\alpha)}{\beta - \alpha} = \frac{\sigma(\alpha) - \sigma(\beta)}{\alpha - \beta}.$$

So we get :

$$\mathrm{sgn}(\sigma) = \prod_{\alpha<\beta} \frac{\sigma(\alpha) - \sigma(\beta)}{\alpha - \beta} = \prod_{i<j} \frac{\sigma(\tau(i)) - \sigma(\tau(j))}{\tau(i) - \tau(j)}.$$

□

Remark 19.12 In particular, for every $\sigma \in \mathfrak{S}_n$, we have $\mathrm{sgn}(\sigma^{-1}) = \mathrm{sgn}(\sigma)^{-1}$. But as $\mathrm{sgn}(\sigma) \in \{\pm 1\}$, this gives :

$$\mathrm{sgn}(\sigma^{-1}) = \mathrm{sgn}(\sigma).$$

3/14/2017 : Snow day

3/16/2017

20 Determinants 3 : alternating n-linear forms on a n-dimensional vector space

Example 20.1 Suppose that $V = K^2$ and that we use the basis $(\vec{e}_1, \vec{e}_2)$. If f is an alternating bilinear form on K^2 such that $f(\vec{e}_1, \vec{e}_2) = 1$, then we get for any $(x_1, x_2), (y_1, y_2) \in K^2$:

$$f((x_1, x_2), (y_1, y_2)) = f(x_1\vec{e}_1 + x_2\vec{e}_2, y_1\vec{e}_1 + y_2\vec{e}_2) =$$

$$= x_1y_1f(\vec{e}_1, \vec{e}_1) + x_1y_2f(\vec{e}_1, \vec{e}_2) + x_2y_1f(\vec{e}_2, \vec{e}_1) + x_2y_2f(\vec{e}_2, \vec{e}_2) = x_1y_2 - x_2y_1.$$

It is easy to see that this formula indeed defines an alternating bilinear form on K^2.

A similar (but longer) calculation would show that, if f is an alternating 3-linear form on K^3 such that $f(\vec{e}_1, \vec{e}_2, \vec{e}_3) = 1$, then :

$$f((x_1, x_2, x_3), (y_1, y_2, y_3), (z_1, z_2, z_3)) = x_1y_2z_3 - x_1y_3z_2 - x_2y_1z_3 + x_2y_3z_1 + x_3y_1z_2 - x_3y_2z_1.$$

We will now generalize this example to n-dimensional vector spaces.

We fix a K-vector space V. For every $n \geq 1$, we write $\bigwedge^n(V, K)$ for the space of alternating n-linear forms on V (definition 18.4).

Proposition 20.2 *Let $f \in \bigwedge^n(V,K)$.*

(i) If $\vec{v}_1,\dots,\vec{v}_n \in V$ and $\sigma \in \mathfrak{S}_n$, then :

$$f(\vec{v}_{\sigma(1)},\dots,\vec{v}_{\sigma(n)}) = \operatorname{sgn}(\sigma) f(\vec{v}_1,\dots,\vec{v}_n).$$

(ii) Let $\vec{v}_1,\dots,\vec{v}_n \in V$, let $A = (a_{ij}) \in M_n(K)$, and write $\vec{w}_i = \sum_{j=1}^n a_{ji}\vec{v}_j$, $1 \le i \le n$.[15]

Let $f \in \bigwedge^n(V,K)$. Then :

$$f(\vec{w}_1,\dots,\vec{w}_n) = \left(\sum_{\sigma\in\mathfrak{S}_n} \operatorname{sgn}(\sigma)\prod_{i=1}^n a_{i,\sigma(i)}\right) f(\vec{v}_1,\dots,\vec{v}_n) = \left(\sum_{\tau\in\mathfrak{S}_n} \operatorname{sgn}(\tau)\prod_{i=1}^n a_{\tau(i),i}\right) f(\vec{v}_1,\dots,\vec{v}_n).$$

Proof. Let's prove (i). By proposition 19.6, we know that σ is a product of transpositions. Write $\sigma = \tau_1 \dots \tau_r$, where each τ_s is a transposition. We prove (i) by induction on r. If $r = 0$, then σ is the identity of $\{1,\dots,n\}$ and there is nothing to prove. Otherwise, we can write $\sigma = \tau\tau_r$, with τ_r a transposition and (i) known for τ. We write $\tau_r = (ij)$ with $i < j$. Then :

$$f(\vec{v}_{\sigma(1)},\dots,\vec{v}_{\sigma(n)}) = f(\vec{v}_{\tau(1)},\dots,\vec{v}_{\tau(i-1)},\vec{v}_{\tau(j)},\vec{v}_{\tau(i+1)},\dots,\vec{v}_{\tau(j-1)},\vec{v}_{\tau(i)},\vec{v}_{\tau(j+1)},\dots,\vec{v}_{\tau(n)}),$$

which is equal to $-f(\vec{v}_{\tau(1)},\dots,\vec{v}_{\tau(n)})$ by (i) of lemma 18.5. By thye induction hypothesis (that is, the formula of (i) for τ), we get :

$$f(\vec{v}_{\sigma(1)},\dots,\vec{v}_{\sigma(n)}) = -\operatorname{sgn}(\tau) f(\vec{v}_1,\dots,\vec{v}_n).$$

The conclusion then follows from the fact that $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\tau)\operatorname{sgn}(\tau_n) = -\operatorname{sgn}(\tau)$. (Use corollary 19.11 and example 19.9.)

Let's prove (ii). The two formulas in (ii) are equivalent, as we see by doing the change of variable $\tau = \sigma^{-1}$ and using the fact that $\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)$ for every $\sigma \in \mathfrak{S}_n$ (remark 19.12). So we will only prove the second formula.

By the n-linearity of f, we have :

$$\begin{aligned} f(\vec{w}_1,\dots,\vec{w}_n) &= f\left(\sum_{j_1=1}^n a_{j_1,1}\vec{v}_{j_1},\dots,\sum_{j_n=1}^n a_{j_n,n}\vec{v}_{j_n}\right) \\ &= \sum_{j_1=1}^n \cdots \sum_{j_n=1}^n a_{j_1,1}\dots a_{j_n,n} f(\vec{v}_{j_1},\dots,\vec{v}_{j_n}) \end{aligned}$$

By the definition of alternating forms, $f(\vec{v}_{j_1},\dots,\vec{v}_{j_n}) = 0$ unless all the $\vec{v}_{j_r}$ are distinct, which is equivalent to saying that the map $r \mapsto j_r$ is an element of $\mathfrak{S}_n$. So we get :

$$f(\vec{w}_1,\dots,\vec{w}_n) = \sum_{\sigma\in\mathfrak{S}_n} a_{\sigma(1),1}\dots a_{\sigma(n),n} f(\vec{v}_{\sigma(1)},\dots,\vec{v}_{\sigma(n)}).$$

[15] In other words, the matrix B with columns $\vec{v}_1,\dots,\vec{v}_n$ and the matrix C with columns $\vec{w}_1,\dots,\vec{w}_n$ are related by $C = BA$.

By (i), this is equal to :

$$\sum_{\sigma\in\mathfrak{S}_n} \mathrm{sgn}(\sigma) a_{\sigma(1),1}\dots a_{\sigma(n),n} f(\vec{v}_1,\dots,\vec{v}_n).$$

□

Theorem 20.3 *Suppose that* $\dim(V) = n < +\infty$. *Then* $\dim(\bigwedge^n(V,K)) = 1$.

More precisely, let $(\vec{v}_1,\dots,\vec{v}_n)$ *be a basis of* V. *Then there exists a unique alternating* n*-linear form* f *on* V *such that* $f(\vec{v}_1,\dots,\vec{v}_n) = 1$, *and every alternating* n*-linear form on* V *is a multiple of* f *(that is, of the form* λf *with* $\lambda \in K$*).*

The unique alternating n-linear form on V sending $(\vec{v}_1,\dots,\vec{v}_n)$ to 1 will be called $\det_{(\vec{v}_1,\dots,\vec{v}_n)}$.

Proof. Let $(\vec{v}_1^*,\dots,\vec{v}_n^*)$ be the dual basis of $(\vec{v}_1,\dots,\vec{v}_n)$ (definition 17.8). We set :

$$f = \sum_{\sigma\in\mathfrak{S}_n} \mathrm{sgn}(\sigma)\vec{v}^*_{\sigma(1)} \otimes \dots \otimes \vec{v}^*_{\sigma(n)},$$

with the definition of (4) of example 18.3. Remember that this just means that, for all $\vec{w}_1,\dots,\vec{w}_n \in V$:

$$f(\vec{w}_1,\dots,\vec{w}_n) = \sum_{\sigma\in\mathfrak{S}_n} \mathrm{sgn}(\sigma)\vec{v}^*_{\sigma(1)}(\vec{w}_1)\dots\vec{v}^*_{\sigma(n)}(\vec{w}_n) = \sum_{\sigma\in\mathfrak{S}_n} \mathrm{sgn}(\sigma)\vec{v}^*_1(\vec{w}_{\sigma(1)})\dots\vec{v}^*_n(\vec{w}_{\sigma(n)}).$$

(We get the second formula by doing the change of variable $\sigma \mapsto \sigma^{-1}$ and using the fact that $\mathrm{sgn}(\sigma^{-1}) = \mathrm{sgn}(\sigma)$.) This is a n-linear form, and, by definition of the $\vec{v}_i^*$, we have $f(\vec{v}_1,\dots,\vec{v}_n) = 1$.

So to prove the theorem, we need to prove the two following facts :

(A) f is alternating.

(B) Every element of $\bigwedge^n(V,K)$ is a multiple of f. (That is, f is a basis of $\bigwedge^n(V,K)$.

Let's prove (A). Let $\vec{w}_1,\dots,\vec{w}_n \in V$, and suppose that we have $\vec{w}_i = \vec{w}_j$ with $1 \le i < j \le n$. Let $\tau = (ij) \in \mathfrak{S}_n$. Then, for every $\sigma \in \mathfrak{S}_n$,

$$(\vec{w}_{\sigma(1)},\dots,\vec{w}_{\sigma(n)}) = (\vec{w}_{\tau\sigma(1)},\dots,\vec{w}_{\tau\sigma(n)}).$$

Let $X = \{\sigma \in \mathfrak{S}_n | \sigma(i) < \sigma(j)\}$. Then, for every $\sigma \in \mathfrak{S}_n$, we have either $\sigma \in X$, or $\tau\sigma = \tau^{-1}\sigma \in X$. So :

$$\begin{aligned} f(\vec{w}_1,\dots,\vec{w}_n) &= \sum_{\sigma\in\mathfrak{S}_n} \mathrm{sgn}(\sigma)\vec{v}^*_1(\vec{w}_{\sigma(1)})\dots\vec{v}^*_n(\vec{w}_{\sigma(n)}) \\ &= \sum_{\sigma\in X} (\mathrm{sgn}(\sigma)\vec{v}^*_1(\vec{w}_{\sigma(1)})\dots\vec{v}^*_n(\vec{w}_{\sigma(n)}) + \mathrm{sgn}(\tau\sigma)\vec{v}^*_1(\vec{w}_{\tau\sigma(1)})\dots\vec{v}^*_n(\vec{w}_{\tau\sigma(n)}) \\ &= \sum_{\sigma\in X} (1+\mathrm{sgn}(\tau))\,\mathrm{sgn}(\sigma)\vec{v}^*_1(\vec{w}_{\sigma(1)})\dots\vec{v}^*_n(\vec{w}_{\sigma(n)}). \end{aligned}$$

As $\mathrm{sgn}(\tau) = 0$, this is equal to 0, and so f is alternating.

Let's prove (B). Take $g \in \bigwedge^n(V, K)$, and let $\lambda = g(\vec{v}_1, \ldots, \vec{v}_n)$. We'll show that $g = \lambda f$. Let $\vec{w}_1, \ldots, \vec{w}_n \in V$. As $(\vec{v}_1, \ldots, \vec{v}_n)$ is a basis of V, we can write $\vec{w}_i = \sum_{j=1}^n a_{ji}\vec{v}_j$, for $1 \leq i \leq n$. Then, by (ii) of proposition 20.2 :

$$f(\vec{w}_1, \ldots, \vec{w}_n) = \left(\sum_{\sigma \in \mathfrak{S}_n} \mathrm{sgn}(\sigma) \prod_{i=1}^n a_{i,\sigma(i)} \right) f(\vec{v}_1, \ldots, \vec{v}_n) = \sum_{\sigma \in \mathfrak{S}_n} \mathrm{sgn}(\sigma) \prod_{i=1}^n a_{i,\sigma(i)}.$$

and

$$g(\vec{w}_1, \ldots, \vec{w}_n) = \left(\sum_{\sigma \in \mathfrak{S}_n} \mathrm{sgn}(\sigma) \prod_{i=1}^n a_{i,\sigma(i)} \right) g(\vec{v}_1, \ldots, \vec{v}_n) = \lambda \sum_{\sigma \in \mathfrak{S}_n} \mathrm{sgn}(\sigma) \prod_{i=1}^n a_{i,\sigma(i)}.$$

So $g(\vec{w}_1, \ldots, \vec{w}_n) = \lambda f(\vec{w}_1, \ldots, \vec{w}_n)$, as desired.

□

21 Determinants 4 : determinant of an endomorphism

In this section, we fix a finite-dimensional K-vector space V, and we let $n = \dim(V)$.

Theorem 21.1 *Let $u \in \mathrm{End}(V)$.*[16] *Then there exists a unique scalar $\det(u) \in K$ such that, for every alternating n-linear form f on K and for all $\vec{v}_1, \ldots, \vec{v}_n$, we have :*

$$f(u(\vec{v}_1, \ldots, \vec{v}_n)) = \det(u) f(\vec{v}_1, \ldots, \vec{v}_n).$$

We call this scalar the determinant *of u.*

Moreover, we have the following properties :

(i) $\det(\mathrm{id}_V) = 1$.

(ii) If $u, v \in \mathrm{End}(V)$, then $\det(u \circ v) = \det(u)\det(v)$.

(iii) $\det(u) \neq 0$ *if and only if u is invertible.*

(iv) If $(\vec{v}_1, \ldots, \vec{v}_n)$ is a basis of V and A is the matrix of u in the basis $(\vec{v}_1, \ldots, \vec{v}_n)$, then

$$\det(u) = \sum_{\sigma \in \mathfrak{S}_n} \mathrm{sgn}(\sigma) \prod_{i=1}^n a_{i,\sigma(i)}.$$

Proof. Remember that we wrote $\bigwedge^n(V, K)$ for the space of alternating n-linear forms on V. Let $u \in \mathrm{End}(V)$. If $f \in \bigwedge^n(V, K)$, we define $u^*(f) : V^n \to K$ by :

$$u^*(f)(\vec{v}_1, \ldots, \vec{v}_n) = f(u(\vec{v}_1), \ldots, u(\vec{v}_n)).$$

[16] This means that u is a linear transformation from V to itself.

As u is linear, this is also a n-linear form, and it is clearly alternating. So $u^*(f) \in \bigwedge^n(V,K)$. Also, it follows readily from the definition that the map $u^* : \bigwedge^n(V,K) \to \bigwedge^n(V,K)$ is K-linear. (This means that $u^*(f+g) = u^*(f)+u^*(g)$ and $u^*(\lambda f) = \lambda u^*(f)$, which is easy to check on the formula for u^*.)

Now we use theorem 20.3. It says that the K-vector space $\bigwedge^n(V,K)$ is of dimension 1, and so any endomorphism of this vector space is of the form $f \mapsto \lambda f$ for some uniquely determined $\lambda \in K$. This means that there exists a unique $\det(u) \in K$ such that $u^*(f) = \det(u)f$ for every $f \in \bigwedge^n(V,K)$. Looking at the definition of u^*, we see that this is exactly the property that we wanted for $\det(u)$.

Now let's prove (i), (ii) and (iii). If $u = \mathrm{id}_V$, then u^* is the identity map of $\bigwedge^n(V,K)$, and so $\det(u) = 1$. For (ii), let $u, v \in \mathrm{End}(V)$. Then, for every $f \in \bigwedge^n(V,K)$,

$$((u\circ v)^*(f))(\vec{v}_1,\dots,\vec{v}_n) = f(uv(\vec{v}_1,\dots,\vec{v}_n)) = (u^*(f))(v(\vec{v}_1,),\dots,\vec{v}_n) = (v^*(u^*f))(\vec{v}_1,\dots,\vec{v}_n),$$

so $(u\circ v)^* = v^*\circ u^*$. As $(u\circ v)^*$ is multiplication by $\det(u\circ v)$ and $v^*\circ u^*$ is multiplication by $\det(v)\det(u)$, this gives (ii).

Let's prove (iii). If u is invertible, then, by (i) and (ii),

$$1 = \det(\mathrm{id}_V) = \det(u\circ u^{-1}) = \det(u)\det(u)^{-1},$$

so $\det(u) \neq 0$. Conversely, suppose that $\det(u) \neq 0$. Let $(\vec{v}_1,\dots,\vec{v}_n)$ be a basis of V, and take $f \in \bigwedge^n(V,K)$ such that $f(\vec{v}_1,\dots,\vec{v}_n) = 1$ (such a f exists by theorem 20.3). Then

$$f(u(\vec{v}_1,\dots,\vec{v}_n) = \det(u)f(\vec{v}_1,\dots,\vec{v}_n) = \det(u) \neq 0,$$

so, by (ii) of lemma 18.5, the vectors $u(\vec{v}_1),\dots,u(\vec{v}_n)$ are linearly independent. By (v) of theorem 9.5, they form a basis of V. As $\mathrm{Im}(u)$ is the span of $u(\vec{v}_1),\dots,u(\vec{v}_n)$, it is equal to V, and so u is an isomorphism (by corollary 16.2).

Finally, (iv) follows from (ii) of proposition 20.2.

□

3/28/2017

22 Determinants 5 : determinant of a square matrix

Fix $n \geq 1$.

Definition 22.1 If $A \in M_n(K)$, the *determinant* of A is by definition the determinant of the linear transformation $K^n \to K^n$ sending $\vec{v}$ to $A\vec{v}$. (In other words, the endomorphism of K^n whose matrix in the canonical basis is A.)

We can use the results of the previous sections to get the expected properties of the determinants.

Proposition 22.2 *Let $A, B \in M_n(K)$.*

(i) $\det(I_n) = 1$.

(ii) $\det(A) \neq 0$ *if and only if* A *is invertible.*

(iii) $\det(AB) = \det(A)\det(B)$.

(iv) *If* B *is invertible, then* $\det(B^{-1}) = \det(B)^{-1}$ *and* $\det(BAB^{-1}) = \det(A)$.

(v) *If* $A = (a_{ij})$, *then :*

$$\det(A) = \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)} = \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{\sigma(i),i}.$$

(vi) *If* V *is a* n*-dimensional* K*-vector space,* $T \in \operatorname{End}(V)$, $\mathfrak{B}$ *is* any *basis of* V *and* A *is the matrix of* T *in* $\mathfrak{B}$, *then* $\det(T) = \det(A)$.

In particular, if $n = 2$, we recover the usual formula :

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & 2_{22} \end{pmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Proof. Remember that the endomorphism u of K^n corresponding to $A \in M_n(K)$ is given by $u(\vec{v}) = A\vec{v}$, that $u = \mathrm{id}_V$ if $A = I_n$, that u is an isomorphism if and only A is invertible and that this bijection $M_n(K) \xrightarrow{\sim} \operatorname{End}(K^n)$ sends matrix multiplication to composition. So (i)-(iii) just follow from theorem 21.1, and (iv) follows from (i) and (iii).

As for (v) and (vi), they both follow immediately from (iv) of theorem 21.1.

□

Here are a few additional properties of determinants of matrices :

Proposition 22.3 *(i)* $\det(A) = \det(A^T)$.

(ii) $\det(A)$ *is linear in the columns and in the rows of* A.

(iii) *If* $\lambda \in K$, *then* $\det(\lambda A) = \lambda^n \det(A)$.

(iv) *If we apply a permutation* σ *to the columns (or the rows) of* A, *then this multiplies* $\det(A)$ *by* $\operatorname{sgn}(\sigma)$. *In particular, switching two columns (or two rows) of* A *multiplies the determinant by* -1.

(v) *A bit more generally, the map that sends* $(\vec{v}_1, \ldots, \vec{v}_n) \in (K^n)^n$ *to the determinant of the matrix with columns* $\vec{v}_1, \ldots, \vec{v}_n$ *is a* n*-linear alternating form. (And we have the same statement if we replace "columns" by "rows".)*

(vi) *If we add a linear combination of columns of* A *to another column of* A, *this does not change the determinant. Same statement for rows.*

(vii) If $A = \begin{pmatrix} a_{11} & & & * \\ & a_{22} & & \\ & & \ddots & \\ 0 & & & a_{nn} \end{pmatrix}$*, that is, all the entries of* A *below the diagonal are* 0 *(such a* A *is called* upper triangular*), then* $\det(A) = a_{11}a_{22}\dots a_{nn}$.

This suggests a method for calculating $\det(A)$ (other than the formula of proposition 22.2(v), which is almost always unwieldy in practice) : Put A in upper triangular form using elementary row operations (we know exactly what that will do to the determinant by (ii), (iv) and (v)), and then apply (vi).

Proof. To prove (i), we use the explicit formula of proposition 22.2(v). Write $A = (a_{ij})$. Then :

$$\det(A) = \sum_{\sigma\in\mathfrak{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}$$

and

$$\det(A^T) = \sum_{\sigma\in\mathfrak{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{\sigma(i),i} = \sum_{\sigma\in\mathfrak{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{\sigma^{-1}(i),i}$$

(we get the last equality by the change of variable $\sigma \mapsto \sigma^{-1}$, using the fact that $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$). For every $\sigma \in \mathfrak{S}_n$, σ is a bijection, so

$$\prod_{i=1}^{n} a_{\sigma^{-1}(i),i} = \prod_{i=1}^{n} a_{s,\sigma(i)}.$$

Finally, we get :

$$\det(A^T) = \sum_{\sigma\in\mathfrak{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)} = \det(A).$$

We can deduce (ii), (iv), (v) and (vi) from the explicit formula of proposition 22.2(v) or from the definition of the determinant. Let's do the second. Let $(\vec{e}_1, \dots, \vec{e}_n)$ be the canonical basis of K^n. By theorem 20.3, there exists a unique alternating n-form f on K^n such that $f(\vec{e}_1, \dots, \vec{e}_n) = 1$. By definition 22.1 and theorem 21.1, $\det(A)$ is the scalar $f(A\vec{e}_1, \dots, A\vec{e}_n)$. But $A\vec{e}_1, \dots, A\vec{e}_n$ are the columns of A, and f is n-linear, so indeed $\det(A)$ is linear and alternating in the columns of A. We also get (iv) by applying proposition 20.2(i) and (v) by applying lemma 18.5(ii). To deduce the statement for the rows, use the fact, proved in (i), that $\det(A) = \det(A^T)$.

(iii) follows from (ii) : We get λA by multiplying each column of A by λ. As the determinant is linear in each column, this multiplies it by λ^n.

To prove (vi), write $A = (a_{ij})$, and use the formula

$$\det(A) = \sum_{\sigma\in\mathfrak{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}.$$

Let $\sigma \in \mathfrak{S}_n$. Suppose that $\sigma \neq 1$. Let i be the biggest element of $\{1, \ldots, n\}$ such that $\sigma(i) \neq i$. Then $\sigma(j) = j$ for $j > n$, and in particular $\sigma(i) \notin \{i+1, \ldots, n\}$, so $\sigma(i) < i$. As A is upper triangular, $a_{i,\sigma(i)} = 0$, and so $\prod_{j=1}^n a_{j,\sigma(j)} = 0$. So the only nonzero term in the formula above for $\det(A)$ is the term for $\sigma = 1$, and we get (vi).

□

23 Determinants 6 : further properties of determinants

In this section, we'll see an additional method to calculate a determinant (very useful in practice), and applications of determinants to calculating the inverse of a matrix and solving linear system (less useful in practice).

Let $A = (a_{ij})$ be a $n \times n$ matrix. For every $i, j \in \{1, \ldots, n\}$, let $A(i,j)$ be the $(n-1) \times (n-1)$ matrix that we obtain by erasing the ith row and the jth column of A.[17]

Theorem 23.1 *(i) For every $i \in \{1, \ldots, n\}$,*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{i,j} \det(A(i,j)).$$

(ii) For every $j \in \{1, \ldots, n\}$,

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{i,j} \det(A(i,j)).$$

The formula in (i) is called "expanding along the ith row" and the formula in (ii) is called "expanding along the jth column".

Proof. If we prove (i), then (ii) will follow by the fact that $\det(A) = \det(A^T)$ (and by aplplying (i) to A^T). So let's prove (i).

Fix $i \in \{1, \ldots, n\}$. To prove (i), we only need to prove that $\sum_{j=1}^n (-1)^{i+j} a_{i,j} \det(A(i,j))$ is linear and alternating in the columns of A, and that it is equal to 1 if $A = I_n$. (This is enough by theorem 20.3 and propositions 22.2 and 22.3.)

Let's first check the second statement. Suppose that $A = I_n$. If $j \neq i$, then the $(n-1) \times (n-1)$ matrix $A(i,j)$ only has $n-2$ nonzero entries, so its determinant is 0 (because formula (v) of proposition 22.2 for $\det(A(i,j))$ only involves products of $n-1$ distinct entries of $A(i,j)$, and each of these products has to be zero). If $i = j$, then $A(i,i) = I_{n-1}$. So

$$\sum_{j=1}^{n} (-1)^{i+j} a_{i,j} \det(A(i,j)) = (-1)^{i+i} a_{i,i} \det(I_{n-1}) = 1.$$

[17] This is not standard notation, because there is no standard notation.

Now we check that $\sum_{j=1}^{n}(-1)^{i+j}a_{i,j}\det(A(i,j))$ is linear and alternating in the columns of A. The linearity is a simple consequence of the formula, so we start with that. We just need to check that each term in the sum is linear in the columns of A. Fix j. Whatever $\det(A(i,j))$ is, we know, by the formula of (v) of proposition 22.2, that it is a sum with some signs of terms of the form $\prod_{r\neq i}a_{r,\sigma(r)}$, with σ a bijection from $\{1,\dots,i-1,i+1,\dots,n\}$ to $\{1,\dots,j-1,j+1,\dots,n\}$. So $a_{ij}\det(A_{ij})$ is linear in each column of A.

Now let's check that $\sum_{j=1}^{n}(-1)^{i+j}a_{i,j}\det(A(i,j))$ is alternating in the columns of A. Let $1\leq\alpha<\beta\leq n$, and suppose that $a_{r,\alpha}=a_{r,\beta}$ for every $r\in\{1,\dots,n\}$. We want to show that $\sum_{j=1}^{n}(-1)^{i+j}a_{i,j}\det(A(i,j))=0$. Let $j\in\{1,\dots,n\}$. If $j\neq\alpha$ and $j\neq\beta$, then $A_{i,j}$ has two identical columns, and so $\det(A_{i,j})=0$. Also, $A_{i,\alpha}$ is just $A_{i,\beta}$ where we moved column number α to the $(\beta-1)$st place, so (by (iv) of proposition 22.3) $\det(A_{i,\alpha})=(-1)^{\beta-1-\alpha}\det(A_{i,\beta})$. Finally, we get :

$$\sum_{j=1}^{n}(-1)^{i+j}a_{i,j}\det(A(i,j))=(-1)^{i+\alpha}\det(A_{i,\alpha})+(-1)^{i+\beta}\det(A_{i,\beta})$$

$$=(-1)^{i+\alpha+\beta-1-\alpha}\det(A_{i,\beta})+(-1)^{i+\beta}\det(A_{i,\beta})=0.$$

□

3/30/2017

Corollary 23.2 *Suppose that the $n\times n$ matrix A can be written :*

$$A=\left(\begin{array}{c|c}B & C\\ \hline 0 & D\end{array}\right)$$

where B is a $r\times r$ matrix, D is a $(n-r)\times(n-r)$ matrix and C is a $r\times(n-r)$ matrix. Then :

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

Proof. Note that we have $A=A_1A_2$, where

$$A_1=\left(\begin{array}{c|c}I_r & 0\\ \hline 0 & D\end{array}\right)$$

and

$$A_2=\left(\begin{array}{c|c}B & C\\ \hline 0 & I_{n-r}\end{array}\right).$$

So $\det(A)=\det(A_1)\det(A_2)$, and we only need to prove that $\det(A_1)=\det(D)$ and $\det(A_2)=\det(B)$. That is, we have reduced the problem to the case where either B or D is the identity matrix.

Both calculations are similar, one uses expanding along columns and the other expanding along rows. Let's start with $\det(A_1)$. We prove by induction on r that, if $A_1 = \left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & D \end{array}\right)$, then $\det(A_1) = \det(D)$. This is clear if $r = 0$ (because then $A_1 = D$). Let's suppose that $r \geq 1$, and let's use expansion along the first column. We get

$$\det(A_1) = (-1)^{1+1} \det(A_1(1,1)),$$

with

$$A_1(1,1) = \left(\begin{array}{c|c} I_{r-1} & 0 \\ \hline 0 & D \end{array}\right).$$

So the induction hypothesis gives $\det(A_1) = \det(A_1(1,1)) = \det(D)$.

Now let's calculate $\det(A_2)$. We prove by induction on $n-r$ that, if $A_2 = \left(\begin{array}{c|c} B & C \\ \hline 0 & I_{n-r} \end{array}\right)$, then $\det(A_2) = \det(B)$. This is clear if $n - r = 0$ (because then $A_2 = B$). Let's suppose that $n - r \geq 1$, and let's use expansion along the last row. We get :

$$\det(A_2) = (-1)^{n+n} \det(A_2(n,n)),$$

with

$$A_2(n,n) = \left(\begin{array}{c|c} B & C \\ \hline 0 & I_{n-r-1} \end{array}\right).$$

So the induction hypothesis gives $\det(A_2) = \det(A_2(n,n)) = \det(B)$.

□

Example 23.3 Let $a_1, \ldots, a_n \in K$. The *Vandermonde determinant* $VdM(a_1, \ldots, a_n)$ is the determinant of the following $n \times n$ matrix (called the *Vandermonde matrix* for $a_1, \ldots, a_n$) :

$$A = \begin{pmatrix} 1 & a_1 & a_1^2 & \ldots & a_1^{n-1} \\ 1 & a_2 & a_2^2 & \ldots & a_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & a_n & a_n^2 & \ldots & a_n^{n-1} \end{pmatrix}$$

I claim that we have :

$$VdM(a_1, \ldots, a_n) = \prod_{1 \leq i < j \leq n} (a_j - a_i).$$

Let's prove this. The method is very similar to the one used in question 10(c) of problem set, and explains the result of this question. We can use the elementary row operations of the solution of PS1 10(c), or use elementary column operations. Let's do the second. First we subtract a_1 times column $n-1$ from column n, then subtract a_1 times column $n-2$ from column $n-1$,. . . , finally subtract a_1 times column 1 from

column 2. These operations don't change the determinant, and in the end we get the matrix :

$$B = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & a_2 - a_1 & a_2(a_2 - a_1) & \dots & a_2^{n-2}(a_2 - a_1) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & a_n - a_1 & a_n(a_n - a_1) & \dots & a_n^{n-2}(a_n - a_1) \end{pmatrix}$$

Expanding along the first row, we get :

$$VdM(a_1, \dots, a_n) = \det(B(1,1)).$$

Using the linearity in each column of the determinant, we get :

$$\det(B(1,1)) = (a_2 - a_1)(a_3 - a_1) \dots (a_n - a_1) VdM(a_2, \dots, a_{n-1}),$$

hence :

$$VdM(a_1, \dots, a_n) = \left(\prod_{i=2}^{n} (a_i - a_1) \right) VdM(a_2, \dots, a_{n-1}).$$

This suggests an induction. If $n = 1$, the result is obvious. If $n \geq 2$ and the result is known for $n - 1$, then the result for n follows directly from the formula above.

Example 23.4 Here is another way to calculate the Vandermonde determinant $VdM(a_1, \dots, a_n)$ of the matrix A defined in example 23.3. First, if two of the a_i are equal, then $VdM(a_1, \dots, a_n) = 0$, because it's the determinant of a matrix with two identical columns. So we may assume that $a_a, \dots, a_n$ are distinct.

Let's make a_1 an indeterminate T and write $P(T) = VdM(T, a_2, \dots, a_n)$. Using the expansion along the first row to calculate the determinant of A gives :

$$P(T) = \sum_{j=1}^{n} (-1)^{j+1} T^{j-1} \det(A(1,j)).$$

In particular, $P(T)$ is a polynomial in T of degree at most $n - 1$. Also, we have $VdM(a_i, a_2, \dots, a_n) = 0$ for every $2 \leq i \leq n$ (because the determinant of a matrix with two identical columns is zero), so $P(a_i) = 0$ for $2 \leq i \leq n$. This means that $P(T)$ is divisible by $(T - a_i)$ for every $i \geq 2$. As the a_i are distinct and $P(T)$ is of degree at most $n - 1$, we get $P(T) = c(T - a_2) \dots (T - a_n)$, for some $c \in K$. This c is just the coefficient of T^{n-1} in $P(T)$, which is $(-1)^{n+1} \det(A(1,n))$ by the formula for $P(T)$ above. By looking at A, we see that $A(1,n)$ is the Vandermonde matrix for $a_2, \dots, a_{n-1}$. So we get :

$$P(T) = (-1)^{n+1}(T - a_2) \dots (T_{a_n}) VdM(a_2, \dots, a_n) = (a_2 - T) \dots (a_n - T) VdM(a_2, \dots, a_n)$$

(because $(-1)^{n+1} = (-1)^{n-1}$). Evaluating at $T = a_1$ gives :

$$VdM(a_1, \dots, a_n) = \left(\prod_{i=2}^{n} (a_i - a_1) \right) VdM(a_2, \dots, a_{n-1}),$$

and then we can finish by induction as in example 23.3.

Definition 23.5 Let $A \in M_n(K)$. The *comatrix* of A (or *matrix of cofactors* of A) is the $n \times n$ matrix $co(A)$ defined by :

$$co(A)_{i,j} = (-1)^{i+j} \det(A(i,j)).$$

In this country, it is more common to talk about the *adjoint matrix* of A, denoted by $\mathrm{adj}(A)$ and defined by $\mathrm{adj}(A) = \mathrm{co}(A)^T$. In other words, we have :

$$\mathrm{adj}(A)_{i,j} = (-1)^{i+j} \det(A(j,i)).$$

The significance of the adjoint matrix comes from the following theorem :

Theorem 23.6 *Let $A \in M_n(K)$. Then we have :*

$$A\,\mathrm{adj}(A) = \mathrm{adj}(A)A = \det(A)I_n.$$

In particular, if A is invertible, then :

$$A^{-1} = \frac{1}{\det(A)}\,\mathrm{adj}(A).$$

Remark 23.7 For 2×2 matrices, this is a well-known formula and you might have seen it already. Write $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then :

$$\mathrm{adj}(A) = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

The matrix A is invertible if and only $\det(A) = ad - bc \neq 0$, and then we have :

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

Now let's prove the theorem.

Proof. Let's calculate $A\,\mathrm{adj}(A)$. Its entry in position (i,j) is (by definition of the matrix product and of $\mathrm{adj}(A)$) :

$$\sum_{r=1}^{n} a_{ik}(-1)^{k+j} \det(A(j,k)).$$

If $j = i$, this is equal to $\det(A)$ by (i) of theorem 23.1 (it's the expansion of $\det(A)$ along the ith row of A). If $j \neq i$, we can also use (i) of theorem 23.1 to identify the expression above : It's the expansion along the ith row of the determinant of the matrix that we get from A by repeating its ith row and deleting its jth row. This matrix has two equal rows, hence its deterinant is 0, and we get :

$$\sum_{r=1}^{n} a_{ik}(-1)^{k+j} \det(A(j,k)) = 0$$

if $j \neq i$. This gives $A\,\mathrm{adj}(A) = \det(A)I_n$. The proof of $\mathrm{adj}(A)A = \det(A)I_n$ is similar, using expansion along columns instead of rows (i.e., (ii) of 23.1).

□

Example 23.8 As you may have guessed, using the adjoint matrix is not the best way to calculate an inverse. (We have already seen a better way before, which is the Gauss algorithm, i.e. using elementary row operations.) However, it does have interesting uses. For example, let $A \in M_n(\mathbb{Q})$ be invertible, and suppose that all the entries of A are in $\mathbb{Z}$. I can ask the following question : When does A^{-1} have all its entries in $\mathbb{Z}$ too ?[18]

Here is the answer : Let A be as above. Then A^{-1} has all its entries in $\mathbb{Z}$ if and only if $\det(A) = \pm 1$.

And here is the proof : First suppose that A^{-1} has all its entries in $\mathbb{Z}$. Then $\det(A^{-1}) \in \mathbb{Z}$, and of course $\det(A) \in \mathbb{Z}$. As $\det(A)\det(A^{-1}) = 1$, this forces $\det(A)$ to be ± 1.

Conversely, suppose that $\det(A) = \pm 1$. As A has all its entries in $\mathbb{Z}$, so does $\mathrm{adj}(A)$. As $A^{-1} = \det(A)^{-1}\,\mathrm{adj}(A)$ and $\det(A)^{-1} = \pm 1$, we see that A^{-1} has all its entries in $\mathbb{Z}$.

From the theorem above, we can deduced Cramer's rule to solve systems of linear equations. (This is not the most general formulation, but it's general enough for us.)

Theorem 23.9 *Let $A \in M_n(K)$, and let $\vec{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \in M_{n1}(K)$ be a column vector. For every $i \in \{1, \ldots, n\}$, we write $A(i|\vec{b})$ for the matrix that we obtain by replacing the ith column of A by $\vec{b}$.*

Suppose that A is invertible. *Then the unique solution of the system $A\vec{x} = \vec{b}$ is given by $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, with :*

$$x_i = \frac{\det(A(i|\vec{b}))}{\det(A)}.$$

Needless to say, this is not the most efficient way to solve a system by hand.

Proof. The unique solution of $A\vec{x} = \vec{b}$ is $\vec{x} = A^{-1}\vec{b}$. By theorem 23.6, this is equal to $\frac{1}{\det(A)}\,\mathrm{adj}(A)\vec{b}$, so :

$$x_i = \frac{1}{\det(A)} \sum_{j=1}^{n} (-1)^{i+j} \det(A(j,i)) b_j.$$

This is $\frac{1}{\det(A)}$ times the expansion along the ith column of the determinant of the matrix $A(i|\vec{b})$.

□

[18] This kind of question comes up, for example, when you're writing an exam subject and want a matrix whose inverse is not too difficult to manipulate.

24 Eigenvalues, eigenvectors and eigenspaces

Definition 24.1 Let V be a K-vector space and T be an endomorphism of V. We say that $\lambda \in K$ is an *eigenvalue* of T if $\mathrm{Ker}(T - \lambda \mathrm{id}_V) \neq \{0\}$; in that case, the subspace $\mathrm{Ker}(T - \lambda \mathrm{id}_V)$ of V is called the *eigenspace* of T for the eigenvalue λ. An *eigenvector* of T (for the eigenvalue λ) is a nonzero element of $\mathrm{Ker}(T - \lambda \mathrm{id}_V)$, that is, a nonzero $\vec{v} \in V$ such that $T(\vec{v}) = \lambda \vec{v}$.

Definition 24.2 Let A be a $n \times n$ matrix with entries in K. We say that $\lambda \in K$ is an *eigenvalue* of A if $\mathrm{Ker}(A - \lambda I_n) \neq \{0\}$; in that case, the subspace $\mathrm{Ker}(A - \lambda I_n)$ of K^n is called the *eigenspace* of A for the eigenvalue λ. An *eigenvector* of A (for the eigenvalue λ) is a nonzero element of $\mathrm{Ker}(A - \lambda I_n)$, that is, a nonzero $\vec{v} \in K^n$ such that $A\vec{v} = \lambda \vec{v}$.

Remark 24.3 If T is an endomorphism of K^n, then its eigenvalues/ eigenspaces/eigenspaces coincide with those of its matrix A in the canonical basis of K^n.

Thanks to this remark, every result about eigenthings that we prove for endomorphisms will go over to matrices.

Example 24.4 0 is an eigenvalue of T if and only if T is not injective.

Remark 24.5 In definition 24.1, suppose that V is finite-dimensional. Then, for $\lambda \in K$, the following conditions are equivalent :

(i) λ is an eigenvalue of T.

(ii) $T - \lambda \mathrm{id}_V$ is not injective, i.e. has kernel $\neq \{0\}$.

(iii) $T - \lambda \mathrm{id}_V$ is not surjective, i.e. has rank $< \dim(V)$.

(iv) $T - \lambda \mathrm{id}_V$ is not an isomorphism, i.e. is not invertible.

(v) $\det(T - \lambda \mathrm{id}_V) = 0$.

Remark 24.6 For a matrix $A \in M_n(K)$, the analogue of the previous remark is the following : For $\lambda \in K$, the following conditions are equivalent :

(i) λ is an eigenvalue of A.

(ii) $Ker(A - \lambda I_n) \neq \{0\}$, i.e. the system $(A - \lambda I_n)\vec{x} = 0$ admits nontrivial solutions.

(iii) $\mathrm{rank}(A - \lambda I_n) < n$.

(iv) $A - \lambda I_n$ is not invertible.

(v) $\det(A - \lambda I_n) = 0$.

These two remarks will motivate the definition of the characteristic polynomial in the next section.

Here is an elementary but fundamental result about eigenspaces.

Theorem 24.7 *Let V be a K-vector space and T an endomorphism of V. Let $\lambda_1, \ldots, \lambda_r \in K$ be* distinct *eigenvalues of T. Then the eigenspaces $\mathrm{Ker}(T-\lambda_1 \mathrm{id}_V) \ldots, \mathrm{Ker}(T-\lambda_r \mathrm{id}_V)$ are in direct sum.*

Proof. Thanks to lemma 10.3, we only to check the following : If $\vec{v}_1 \in \mathrm{Ker}(T-\lambda_1 \mathrm{id}_V), \ldots,$ $\vec{v}_r \in \mathrm{Ker}(T-\lambda_r \mathrm{id}_V)$ are such that $\vec{v}_1 + \cdots + \vec{v}_r = 0$, then $\vec{v}_1 = \cdots = \vec{v}_r = 0$.

So suppose that this is not true. We choose a family $\vec{v}_1 \in \mathrm{Ker}(T-\lambda_1 \mathrm{id}_V), \ldots, \vec{v}_r \in \mathrm{Ker}(T-\lambda_r \mathrm{id}_V)$ such that $\vec{v}_1 + \cdots + \vec{v}_r = 0$, that not all $\vec{v}_i$ are 0, and that the number s of nonzero $\vec{v}_i$ is as small as possible. Let $1 \leq i_1 < \cdots < i_s \leq r$ be the indices i such that $\vec{v}_i \neq 0$. We have $\vec{v}_{i_1} + \cdots + \vec{v}_{i_s} = 0$. (Note that this implies taht $s \geq 2$.) Applying T and using the fact that $\vec{v}_i \in \mathrm{Ker}(T-\lambda_i \mathrm{id}_V)$ for every i gives,

$$\lambda_{i_1} \vec{v}_{i_1} + \cdots + \lambda_{i_s} \vec{v}_{i_s} = 0,$$

while multiplying by λ_{i_1} gives

$$\lambda_{i_1} \vec{v}_{i_1} + \cdots + \lambda_{i_s} \vec{v}_{i_1} = 0.$$

If we substract the second equality from teh first, we get

$$(\lambda_{i_2} - \lambda_{i_1}) \vec{v}_{i_2} + \cdots + (\lambda_{i_s} - \lambda_{i_1}) \vec{v}_{i_s} = 0.$$

Also, for every $j \in \{2, \ldots, s\}$, $(\lambda_{i_j} - \lambda_{i_1}) \vec{v}_{i_j}$ is in $\mathrm{Ker}(T - \lambda_{i_j} \mathrm{id}_V)$ and nonzero (here we use that $\lambda_{i_j} \neq \lambda_{i_1}$). So, setting $\vec{v}'_i = (\lambda_{i_j} - \lambda_{i_1}) \vec{v}_{i_j}$ if $i = i_j$ and $\vec{v}'_i = 0$ if $i \notin \{i_2, \ldots, i_s\}$, we have a found a relation $\vec{v}'_1 + \cdots + \vec{v}'_r = 0$ with each $\vec{v}'_i$ in $\mathrm{Ker}(T - \lambda_i \mathrm{id}_V)$, with some $\vec{v}'_i$ nonzero, and with the number of nonzero $\vec{v}'_i$ smaller than s. This contradicts the choice of the family $\vec{v}_1, \ldots, \vec{v}_r$.

□

25 The characteristic polynomial

If V is finite-dimensional and $T \in \mathrm{End}(V)$, we have seen that $\lambda \in K$ is an eigenvalue of T if and only if $\det(T - \lambda \mathrm{id}_V) = 0$. This motivates the introduction of the following polynomial, as a convenient way to find all the eigenvalues of T.

Definition 25.1 Let $A \in M_n(K)$. The *characteristic polynomial* of A is the polynomial

$$f_T(A) = \det(XI_n - A).$$

Remember also the following definition from PS2 problem 6 and PS5 problem 3 :

Definition 25.2 Let $A = (a_{ij}) \in M_n(K)$. The *trace* of A is the scalar

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

Proposition 25.3 *Let $A \in M_n(K)$.*

(i) If $\lambda \in K$, then λ is an eigenvalue of A if and only if $f_A(\lambda) = 0$.

(ii) $f_A(X)$ is a polynomial of degree n. Its leading term is X^n, the coefficient of X^{n-1} is $-\operatorname{Tr}(A)$ and its constant term is $(-1)^n \det(A)$.

(iii) If $S \in M_n(K)$ is an invertible matrix, then $f_A(X) = f_{SAS^{-1}}(X)$.

Proof. (i) follows from the definition of $f_A(X)$ and the remark above that λ is an eigenvalue of A if and only if $\det(A - \lambda \mathrm{id}_V) = 0$.

Write $A = (a_{ij})$. Then $XI_n - A = (b_{ij})$, with $b_{ij} = -a_{ij}$ if $i \neq j$, and $b_{ii} = X - a_{ii}$. We know that :

$$f_A(X) = \det(XI_n - A) = \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(n) \prod_{i=1}^{n} b_{i\sigma(i)}.$$

Note that the degree of the polynomial $\prod_{i=1}^n b_{i,\sigma(i)}$ is $|\{i \in \{1, \dots, n\} | \sigma(i) = i\}|$. If $\sigma \neq 1$, then there are at most $n-2$ elements $i \in \{1, \dots, n\}$ such that $\sigma(i) = i$, so $\deg(\prod_{i=1}^n b_{i,\sigma(i)}) \leq n-2$. If $\sigma = 1$, then

$$\prod_{i=1}^{n} b_{i,\sigma(i)} = \prod_{i=1}^{n} (X - a_{ii})$$

has degree n and leading term X^n, and the X^{n-1} term in this polynomial is

$$-a_{11}X^{n-1} - a_{22}X^{n-1} - \cdots - a_{nn}X^{n-1} = -\operatorname{Tr}(A)X^{n-1}.$$

This proves that $\deg(f_A(X)) = n$, that the leading term of $f_A(X)$ is X^n, and that the coefficient of X^{n-1} is $-\operatorname{Tr}(A)$. Finally, the constant term of $f_A(X)$ is $f_A(0) = \det(0I_n - A) = \det(-A) = (-1)^n \det(A)$.

For (iii), just note that, is $S \in M_n(K)$ is invertible, then :

$$f_{SAS^{-1}}(X) = \det(XI_n - SAS^{-1}) = \det(S(XI_n - A)S^{-1}) = \det(S)\det(XI_n - A)\det(S)^{-1} = f_A(X).$$

□

4/4/2017

We can now define the characteristic polynomial of an endomorphism.

Proposition-Definition 25.4 *Let V be a finite-dimensional vector space and let T be an endomorphism of V. If $\mathfrak{A}$ is a basis of V and $A = [T]_{\mathfrak{A}}$, we set $f_T(X) = f_A(X)$. This does not depend on the choice of the basis $\mathfrak{A}$ and we call it the* characteristic polynomial *of teh endomorphism T.*

Proof. This is an application of the change of basis formula. Let $\mathfrak{B}$ be another basis of V and let $B = [T]_{\mathfrak{B}}$. By the change of basis formula (proposition 14.7), there exists an invertible matrix S such that $B = SAS^{-1}$.[19] Then we have $f_A(T) = f_B(T)$ by (iii) of proposition 25.3.

□

Applying proposition 25.3 immediately gives the following properties of $f_T(X)$:

Corollary 25.5 *Let T be an endomorphism of a finite-dimensional vector space V. Then :*

(i) If $\lambda \in K$, then λ is an eigenvalue of T if and only if $f_T(\lambda) = 0$.

(ii) $f_T(X)$ is a polynomial of degree n. Its leading term is X^n and its constant term is $(-1)^n \det(T)$.

Remark 25.6 We can also define $\mathrm{Tr}(T)$ to be minus the coefficient of $X^{\dim(V)-1}$ in $f_T(X)$. Then, if $\mathfrak{B}$ is any basis of V and $A = [T]_{\mathfrak{B}}$, we have $\mathrm{Tr}(T) = \mathrm{Tr}(A)$.

Proposition 25.7 *If $T \in \mathrm{End}(V)$ and $\dim V = n$, then T has at most n distinct eigenvalues.*

Proof. Let $\lambda_1, \ldots, \lambda_r$ be distinct eigenvalues of T. For each i, we choose an eigenvector $\vec{v}_i$ with eigenvalue λ_i. By theorem 24.7, the family $(\vec{v}_1, \ldots, \vec{v}_r)$ is linearly independent. So we must have $r \leq n$.

□

Example 25.8 Let $A = (a_{ij}) \in M_n(K)$, and suppose that A is upper triangular, which means that $a_{ij} = 0$ if $i > j$. That is, the entries of A below the diagonal are 0, and we have :

$$A = \begin{pmatrix} a_{11} & & & * \\ & a_{22} & & \\ & & \ddots & \\ 0 & & & a_{nn} \end{pmatrix}$$

and

$$XI_n - A = \begin{pmatrix} X - a_{11} & & & * \\ & X - a_{22} & & \\ & & \ddots & \\ 0 & & & X - a_{nn} \end{pmatrix}.$$

Then we can apply (vii) of propoposition 22.3 to calculate $f_A(X)$. We get :

$$f_A(X) = (X - a_{11})(X - a_{22}) \ldots (X - a_{nn}).$$

[19] Note that we don't care what S is, just that it is invertible.

By proposition 25.3(i), the eigenvalues of A are $a_{11}, \ldots, a_{nn}$.

In particular, $f_{I_n}(X) = (X-1)^n$, and the only eigenvalue of I_n is 1 (which was of course clear from the definition of eigenvalues).

The following property of characteristic polynomials will be very useful in induction arguments later.

Proposition 25.9 *Let V be a finite-dimensional vector space, let T be an endomorphism of V, and let W be a subspace of V such that $T(W) \subset W$. We write T_W for the endomorphism of W given by T. Then $f_{T_W}(X)$ divides $f_T(X)$.*

More precisely, if we choose a basis $\mathfrak{B} = (\vec{v}_1, \ldots, \vec{v}_n)$ of V such that $(\vec{v}_1, \ldots, \vec{v}_r)$ is a basis of W (which we know is always possible), and if $A = [T]_{\mathfrak{B}}$, then the fact that $T(W) \subset W$ implies that we have

$$A = \left(\begin{array}{c|c} B & C \\ \hline 0 & D \end{array}\right),$$

where B is the matrix of T_W in the basis $(\vec{v}_1, \ldots, \vec{v}_r)$, and we have :

$$f_T(X) = f_{T_W}(X) f_D(X).$$

Proof. This follows directly from corollary 23.2.

□

26 Diagonalization

First we introduce some vocabulary about matrices.

Definition 26.1 (1) A square matrix $A = (a_{ij})$ is called *diagonal* if all its non-diagonal entries are 0, that is, if $a_{ij} = 0$ for $i \neq j$.

(2) Two matrices $A, B \in M_n(K)$ are called *similar* if there exists $S \in M_n(K)$ invertible such that $A = SBS^{-1}$.

Remark 26.2 (a) The definition of "similar" is symmetric in A and B (because if $A = SBS^{-1}$, then $B = S^{-1}AS$).

(b) A matrix is always similar to itself (use $S = I_n$).

(c) Suppose that A and B are similar, and that B and C are similar. Then A and C are similar. (Indeed, if $B = SAS^{-1}$ and $C = PBP^{-1}$, then $C = (PS)A(PS)^{-1}$.)

Remark 26.3 If A is similar to I_n, then $A = I_n$. Indeed, let $S \in M_n(K)$ invertible such that $A = SI_nS^{-1}$. Then $A = SS^{-1} = I_n$.

A bit more generally (and with the same proof), if A is similar to λI_n, then $A = \lambda I_n$.

Now we define diagonalizable endomorphisms and matrices.

Definition 26.4 (1) Let V be a finite-dimensional vector space and T be an endomorphism of V. We say that T is *diagonalizable* if there exists a basis $(\vec{v}_1, \ldots, \vec{v}_n)$ such that every $\vec{v}_i$ is an eigenvector of T. (Cf definition 24.1.)

(2) A matrix $A \in M_n(K)$ is called *diagonalizable* if it is similar to a diagonal matrix.

Example 26.5 Let $T : \mathbb{C}^2 \to \mathbb{C}^2$ be the endomorphism whose matrix in the canonical basis is $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. Then, by problem 1 of PS4, T is diagonalizable.

Proposition 26.6 *Let V be a finite-dimensional vector space and T be an endomorphism of V. Then the following statements are equivalent :*

(i) T is diagonalizable.

(ii) There exists a basis $\mathfrak{B}$ of V such that the matrix $[T]_{\mathfrak{B}}$ of T in $\mathfrak{B}$ is diagonal.

(iii) There exists a basis $\mathfrak{B}$ of V such that $[T]_{\mathfrak{B}}$ is diagonalizable.

(iv) For every basis $\mathfrak{B}$ of V, $[T]_{\mathfrak{B}}$ is diagonalizable.

Proof. (ii) is just a reformulation of (i). As a diagonal matrix is diagonalizable (because it's similar to itself), (ii) implies (iii).

We show that (iii) implies (iv). Assume that (iii) is true for a basis $\mathfrak{B}$ of V, and let $\mathfrak{A}$ be another basis of V. Then, by the change of basis formula, there is an invertible matrix $S \in M_n(K)$ such that $[T]_{\mathfrak{A}} = S[T]_{\mathfrak{B}}S^{-1}$. By assumption, $[T]_{\mathfrak{B}}$ is diagonalizable, so there exists a diagonal matrix $D \in M_n(K)$ and an invertible matrix $C \in M_n(K)$ such that $[T]_{\mathfrak{B}} = CDC^{-1}$. Then :

$$[T]_{\mathfrak{A}} = S[T]_{\mathfrak{B}}S^{-1} = S(CDC^{-1})S^{-1} = (SC)D(SC)^{-1},$$

and SC is invertible, so $[T]_{\mathfrak{A}}$ is diagonalizable.

Now let's show that (iv) implies (ii). Let $\mathfrak{A} = (\vec{v}_1, \ldots, \vec{v}_n)$ be a basis of V. By (iv), $[T]_{\mathfrak{A}}$ is diagonalizable, so there exists a diagonal matrix $D \in M_n(K)$ and an invertible matrix $S \in M_n(K)$ such that $[T]_{\mathfrak{A}} = SDS^{-1}$. In other words, $D = S^{-1}[T]_{\mathfrak{A}}S$. Write $S = (c_{ij})$. For every $j \in \{1, \ldots, n\}$, define $\vec{w}_j \in V$ by

$$\vec{w}_j = \sum_{i=1}^{n} c_{ij}\vec{v}_i.$$

In other words, $\vec{w}_j$ is the vector of V with coordinates $[\vec{w}_j]_{\mathfrak{A}}$ in the basis $\mathfrak{A}$ given by the jth column of S.

I claim that $\mathfrak{B} = (\vec{w}_1, \ldots, \vec{w}_n)$ is a basis of V. Indeed, the vectors $\vec{w}_1, \ldots, \vec{w}_n$ have to be linearly indepedent, because otherwise the columns of S would be linearly dependent, which contradicts the fact that S is invertible.

Now that we know that $\mathfrak{B}$ is a basis, we can use the change of basis formula, which says that $[T]_{\mathfrak{B}} = S^{-1}[T]_{\mathfrak{A}}S$. In other words, $[T]_{\mathfrak{B}} = D$, so we have found a basis where the matrix of T is diagonal, and this proves (ii).

□

In general, it is not always easy to say whether an endomorphism (or a matrix) is diagonalizable. But here is a sufficient condition. First, remember the following definition.

Definition 26.7 Let $f(X) \in K[X]$. We say that a root (=zero) $\lambda \in K$ of $f(X)$ is *simple* if $(X - \lambda)^2$ does not divide $f(X)$.

Proposition 26.8 *Let $T \in \mathrm{End}(V)$ as above. If $f_T(X)$ has all its roots in K and all these roots are simple, then T is diagonalizable.*

We have a similar statement for matrices.

Proof. The second statement follows from proposition 26.6 above. Let's prove the first. If $f_T(X)$ has all its roots in K, then we have $f_T(X) = (X - \lambda_1)\dots(X - \lambda_n)$ ($n = \dim V = \deg(f_T(X))$). The fact that all the roots are simple means that the λ_i are all different. For every $i \in \{1, \dots, n\}$, we know that λ_i is an eigenvalue of T by corollary 25.5, so we can pick an eigenvector $\vec{v}_i$ for the eigenvalue λ_i. By theorem 24.7, the vectors $\vec{v}_1, \dots, \vec{v}_n$ are linearly independent. As $\dim V = n$, $(\vec{v}_1, \dots, \vec{v}_n)$ is a basis of V. So T is indeed diagonalizable.

□

Remark 26.9 The condition in the proposition above is far from necessary. For example, I_n is diagonalizable (it's even diagonal), but its characteristic polynomial $f_{I_n}(X) = (X - 1)^n$ doesn't have simple roots.

Example 26.10 Here is a very useful application of the proposition. Let $A \in M_n(K)$, and suppose that A is upper triangular, so that

$$A = \begin{pmatrix} a_{11} & & & * \\ & a_{22} & & \\ & & \ddots & \\ 0 & & & a_{nn} \end{pmatrix}.$$

Suppose also that *the diagonal entries $a_{11}, \dots, a_{nn}$ are all different.* Then A is automatically diagonalizable, and in fact it's similar to the diagonal matrix

$$\begin{pmatrix} a_{11} & & & 0 \\ & a_{22} & & \\ & & \ddots & \\ 0 & & & a_{nn} \end{pmatrix}.$$

Indeed, we have seen that the characteristic polynomial of A is $(X - a_{11}) \dots (X - a_{nn})$.

For example, you can tell just by looking at it that the matrix

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

is diagonalizable and that its eigenvalues are $1, -1, -5$. But this doesn't give you a basis of eigenvectors, you still have to calculate it.

Remark 26.11 The conclusion in the example above becomes totally false if the diagonal entries of A are not distinct. For example, take $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Then A is not diagonalizable.

Indeed, the only eigenvalue of A is 1 by example 25.8. So if A were diagonalizable, then it would be similar to the matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2$. But we have seen that the only matrix similar to I_2 is I_2 itself, and obviously $A \neq I_2$.

Diagonalizing an endomorphism T means finding a basis where the matrix of T is diagonal, and diagonalizing a matrix A means finding an invertible matrix S such that SAS^{-1} is diagonal. The general procedure to diagonalize an endomorphism T is the following (if you want to diagonalize a matrix A, you can apply this procedure to the endomorphism T of K^n with matrix A in the canonical basis) :

(1) Find the eigenvalues of T : The most direct method is to calculate $f_T(X)$ and try to find all its roots. You can also try to solve the equation $T(\vec{v}) - \lambda\vec{v} = 0$, where λ is a parameter, and determine for which values of λ it has nonzero solutions.

(2) For every eigenvalue λ of T, calculate $\mathrm{Ker}(T - \lambda \mathrm{id}_V)$ and find a basis of it.

(3) Put all the bases of step (2) together. If you get a family with $\dim(V)$ elements, you've won. Otherwise, T was not diagonalizable.

4/6/2017

27 Triangularization

Even when T is not diagonalizable, we might still be able to find a basis of V where the matrix of T is upper triangular. This is sometimes called *triangularization* and is much easier.

Definition 27.1 We say that an endomorphism T of a finite-dimensional vector space V is *triangularizable* if there exists a basis of V in which the matrix of T is upper triangular.

We say that a matrix is *triangularizable* if it is similar to an upper triangular matrix.

Remark 27.2 If we replace "upper triangular" by "lower triangular" in the definition above, we pbtain equivalent notions. (Exercise : why ?)

Remark 27.3 In problem 2 of problem set 4, we've seen that the two following statements are equivalent :

(i) Then there exists a basis $\mathfrak{B}$ of V such that $[T]_{\mathfrak{B}}$ is upper triangular (i.e. T is triangularizable).

(ii) There exists a sequence of subspaces $\{0\} = V_0 \subset V_1 \subset \cdots \subset V_n = V$ such that $\dim(V_i) = i$ and $T(V_i) \subset V_i$, for all $0 \leq i \leq n$.

Theorem 27.4 *Let $T \in \mathrm{End}(V)$ as above, and suppose that $f_T(X)$ has all its roots in K.*[20] *Then T is triangularizable.*

We have a similar statement for matrices : If $A \in M_n(K)$ is such that $f_A(X)$ has all its roots in K, then A is triangularizable.

Remark 27.5 The converse is true : If there exists a basis of V in which the matrix of T is upper triangular, then $f_T(X)$ has all its roots in K. Indeed, we have seen in example 26.10 that the characteristic polynomial of an upper triangular matrix is of the form $(X - \lambda_1) \dots (X - \lambda_n)$, where $\lambda_1, \dots, \lambda_n$ are the diagonal entries; such a polyonimal has all its roots in K.

Remark 27.6 By proposition 25.9 and the theorem above, if T is triangularizable and W is a subspace of V such that $T(W) \subset W$, then the endomorphism T_W of W induced by T is also triangularizable.

Proof. The statement for matrices follows from the statement for endomorphisms. Let's prove the statement for endomorphisms. Let $n = \dim V$. We do an induction on n.

If $n = 1$, then T is a multiplication by a scalar, so its matrix is upper triangular in any basis.

Suppose that $n \geq 2$ and that we know the result for $n - 1$. As $f_T(X)$ has all its roots in K, we can pick a root $\lambda_1 \in K$ of $f_T(X)$, which is an eigenvalue of T. Let $\vec{v}_1 \in V$ be an eigenvector of T for the eigenvalue λ_1. Complete $\vec{v}_1$ to a basis $(\vec{v}_1, \dots, \vec{v}_n)$ of V. Then the matrix of T in $\mathfrak{B}$ is of the form

$$\left(\begin{array}{c|c} \lambda_1 & * \\ \hline 0 & B \end{array}\right).$$

Let $W = \mathrm{Span}(\vec{v}_2, \dots, \vec{v}_n)$, and let $p : V \to W$ be the linear transformation defined by $p(\vec{v}_1) = 0$ and $p(\vec{v}_i) = \vec{v}_i$ for $2 \leq i \leq n$. Then the $(n-1) \times (n-1)$ matrix B is the matrix of the endomorphism $p \circ T_{|W}$ of W in the basis $(\vec{v}_2, \dots, \vec{v}_n)$. By proposition 25.9, we have $f_T(X) = (X - \lambda_1) f_B(X)$, so the chaacteristic polynomial $f_B(X)$ of $p \circ T_{|W}$ has all its roots in K. By the induction hypothesis, we can find a basis $(\vec{w}_2, \dots, \vec{w}_n)$ in which

[20] For example, this is automatically true if $K = \mathbb{C}$.

the matrix of $p \circ T_{|W}$ is upper triangular. Then $(\vec{v}_1, \vec{w}_2, \ldots, \vec{w}_n)$ is a basis of V, and the matrix of T in this basis is upper triangular.

□

Example 27.7 The matrix $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ has characteristic polynomial $X^2 + 1$. So it is not similar to an upper triangular matrix as an element of $M_2(\mathbb{R})$, but it is as an element of $M_2(\mathbb{C})$.

Remark 27.8 By proposition 25.9 and

The following lemma will be used again and again.

Lemma 27.9 *Let T and U be two endomorphisms of a K-vector space V, and suppose that $T \circ U = U \circ T$. Let λ be an eignevalue of T, and let $E_\lambda = \mathrm{Ker}(T - \lambda \mathrm{id}_V)$ be the corresponding eigenspace. Then $U(E_\lambda) \subset E_\lambda$.*

Proof. If $\vec{v} \in E_\lambda$, then

$$T(U(\vec{v})) = U(T(\vec{v})) = U(\lambda \vec{v}) = \lambda U(\vec{v}),$$

so $U(\vec{v}) \in E_\lambda$.

□

Theorem 27.10 *(Simultaneous triangularization.) Let T and U be two endomorphism of a finite-dimensional vector space V. Suppose that T and U are both triangularizable and that $T \circ U = U \circ T$. Then there exists a basis $\mathfrak{B}$ of V such that both $[T]_{\mathfrak{B}}$ and $[U]_{\mathfrak{B}}$ are upper triangular.*

Similarly, if A and B are triangularizable matrices in $M_n(K)$ and if $AB = BA$, then there exists $S \in M_n(K)$ invertible such that both SAS^{-1} and SBS^{-1} are upper triangular.

Remark 27.11 There is an analogous result with "diagonalizable" instead of "triangularizable" (and "diagonal" instead of "upper triangular"), but we don't yet have the tools to prove it.

Proof. The statement for matrices follows from the statement for endomorphisms. Let's prove the statement for endomorphisms. Let $n = \dim V$. We do an induction on n.

If $n = 1$, then T and U are both multiplication by some scalar, so their matrices are upper triangular in any basis.

Suppose that $n \geq 2$ and that we know the result for $n - 1$. First we need to show that T and U have a common eigenvector. As $f_T(X)$ has all its roots in K, we can pick a root $\lambda \in K$ of $f_T(X)$, which is an eigenvalue of T. Let $E_\lambda = \mathrm{Ker}(T - \lambda \mathrm{id}_V)$ be the correspondaing eigenspace of T. By lemma 27.9 above, $U(E_\lambda) \subset E_\lambda$. By remark 27.6

above, the restriction of U to E_λ is triangularizable, so U has at least one eigenvector $\vec{v}_1$ in E_λ; let μ be its eigenvalue. As $\vec{v}_1$ is in E_λ, it's also an eigenvector of T with eigenvalue λ.

Complete $\vec{v}_1$ to a basis $(\vec{v}_1, \ldots, \vec{v}_n)$ of V. Then the matrix of T in $\mathfrak{B}$ is of the form

$$\left(\begin{array}{c|c} \lambda & * \\ \hline 0 & B \end{array}\right),$$

and the matrix of U in $\mathfrak{B}$ is of the form

$$\left(\begin{array}{c|c} \mu & * \\ \hline 0 & C \end{array}\right),$$

Let $W = \mathrm{Span}(\vec{v}_2, \ldots, \vec{v}_n)$, and let $p : V \to W$ be the linear transformation defined by $p(\vec{v}_1) = 0$ and $p(\vec{v}_i) = \vec{v}_i$ for $2 \leq i \leq n$. Then the $(n-1) \times (n-1)$ matrix B is the matrix of the endomorphism $p \circ T_{|W}$ of W in the basis $(\vec{v}_2, \ldots, \vec{v}_n)$, and C is the matrix of the endomorphism $p \circ U_{|W}$ of W in the basis $(\vec{v}_2, \ldots, \vec{v}_n)$. By proposition 25.9, we have $f_T(X) = (X - \lambda) f_B(X)$ and $f_U(X) = (X - \mu) f_C(X)$, so the characteristic polynomials $f_B(X)$ and $f_C(X)$ of $p \circ T_{|W}$ and $p \circ U_{|W}$ have all their roots in K. By the induction hypothesis, we can find a basis $(\vec{w}_2, \ldots, \vec{w}_n)$ in which the matrices of $p \circ T_{|W}$ and $p \circ U_{|W}$ are both upper triangular. Then $(\vec{v}_1, \vec{w}_2, \ldots, \vec{w}_n)$ is a basis of V, and the matrices of T and U in this basis and upper triangular.

□

28 Some properties of polynomials

Remember the following definitions for polynomials.

Definition 28.1 (1) If $P, A \in K[X]$, we say that Q *divides* P if there exists $R \in K[X]$ such that $P = QR$. This implies that $\deg(Q) \leq \deg(P)$.

(2) If $P \in K[X]$, we say that $a \in K$ is a *root* (or *zero*) of P if $P(a) = 0$; this is equivalent to saying that $X - a$ divides P. In this case, the *multiplicity* of the root a of P is the biggest integer m such that $(X - a)^m$ divides P, and we say that a is a *simple root* of P if its multiplicity is 1 (i.e. if $(X - a)^2$ does not divide P).

(3) We say that a nonzero polynomial $P \in K[X]$ is *monic* if the coefficient of its highest degree term is 1.

Theorem 28.2 *(Euclidian division for polynomials.) Let $A, B \in K[X]$, and assume that $B \neq 0$. Then there exists a unique pair of polynomials (Q, R) such that :*

(1) $A = BQ + R$;

(2) $\deg(R) < \deg(B)$.

We say that Q (resp. R) is the *quotient* (resp. *remainder*) of the Euclidian division of A by B. Note that $R = 0$ if and only if B divides A.

If you've seen Euclidian division for integers, this is a very similar result.

Proof. Let's first prove uniqueness of Q and R. Suppose that we have two pairs of polynomials (Q_1, R_1) and (Q_2, R_2) satisfying conditions (1) and (2). Then $A = BQ_1 + R_1 = BQ_2 + R_2$, so $B(Q_1 - Q_2) = R_2 - R_1$, hence $\deg(B) + \deg(Q_1 - Q_2) = \deg(R_2 - R_1)$. As R_1 and R_2 have degree $< \deg(B)$, so does $R_2 - R_1$, and the equality above is only possible if $Q_1 - Q_2 = 0$, i.e. $Q_1 = Q_2$. Then this also implies that $R_1 = R_2$.

We now prove existence by induction on $\deg(A)$. If $\deg(A) < \deg(B)$, we take $Q = 0$ and $R = A$. Suppose that $\deg(A) \geq \deg(B)$ and that we know the results for all polynomials A_1 of degree $< \deg(A)$. Let $d = \deg(A)$ and $r = \deg(B) \leq d$. Write $A = \sum_{i=0}^{d} a_i X^i$, $B = \sum_{i=0}^{r} b_i X^i$, and set $A_1 = A - a_d b_r^{-1} X^{d-r} B$. Then $\deg(A_1) < d = \deg(A)$, so by the induction hypothesis we have a couple (Q_1, R_1) of polynomials such that $A_1 = BQ_1 + R_1$ and $\deg(R_1) < \deg(B)$. We can now take $Q = a_d b_r^{-1} X^{d-r} + Q_1$ and $R = R_1$.

□

We will now give the main consequence of this result. We introduce a convenient piece of vocabulary from commutative algebra for this (but you don't need to remember this definition for the homework of exams).

4/11/2017

Definition 28.3 A subset I of $K[X]$ is called an *ideal* if :

(0) $0 \in I$.

(1) For every $A, B \in I$, $A + B \in I$.

(2) For every $A \in I$ and every $P \in K[X]$[21], PA is in I.

Example 28.4 If $A \in K[X]$, then the set I of multiples of A is an ideal of $K[X]$, traditionnally denoted by (A).

The existence of Euclidian division implies that every ideal is of the form above.

Corollary 28.5 *If I is an ideal of $K[X]$ and $I \neq \{0\}$, then there exists a unique monic polynomial $A \in K[X]$ such that $I = (A)$ (i.e. I is the set of multiples of A).*

Proof. First we prove uniqueness. Suppose that we have two monic polynomials A and B such that $I = (A) = (B)$. Then A divides B and B divides A, so there exists $c \in K$ nonzero such that $A = cB$. As both A and B are monic, this implies that $A = B$.

[21] not just I !

Let's prove existence. Let A be a nozero polynomial of minimal degree in I. After multiplying A by a scalar, we may assume that A is monic. By definition of an ideal, we have $(A) \subset I$. Let's prove that $I \subset (A)$. Let $B \in I$. By the theorem above, there exist polynomials $Q, R \in K[X]$ such that $B = AQ + R$ and $\deg(R) < \deg(A)$. As $A, B \in I$, $R = B - QA \in I$. As no nonzero element of I can have degree $\deg(A)$ (by the choice of A), $R = 0$, which means that A divides B, i.e. $B \in (A)$.

□

This result allows us, after defining the greatest common divisor of a family of polynomials, to extend Bezout's theorem to polynomials.

Definition 28.6 Let $P_1, \ldots, P_n \in K[X]$ not all zero. The *greatest common divisor* (gcd) of $P_1, \ldots, P_n$ is of the monic polynomial D of maximal degree that divides all the P_i for $1 \leq i \leq n$.

We say that the polynomials $P_1, \ldots, P_n$ are *relatively prime* if their greatest common divisor is 1.

Theorem 28.7 *(Bezout's theorem.) Let $P_1, \ldots, P_n$ be polynomials that are not all zero, and let D be their greatest common divisor. Then there exist polynomials $Q_1, \ldots, Q_n$ such that $D = P_1Q_1 + \cdots + P_nQ_n$.*

In particular, if $P_1, \ldots, P_n$ are relatively prime, then their exist polynomials $Q_1, \ldots, Q_n$ such that $P_1Q_1 + \cdots + P_nQ_n = 1$.

Proof. Let I be the set of all polynomials of the form $P_1Q_1 + \cdots + P_nQ_n$, for $Q_1, \ldots, Q_n$ varying in $K[X]$. It follows directly from the definition that I is an ideal, and we have $I \neq \{0\}$ because all the P_i are in I and at least of the P_i is nonzero.

By corollary 28.5, there exists a unique monic polynomial $A \in K[X]$ such that $I = (A)$. I claim that $A = D$, which implies the theorem by definition of I. First, since each P_i is in I, A divides each P_i, so by definition of D we have $\deg(A) \leq \deg(D)$. Second, because D divides each P_i, it also divides each polynomial of the form $P_1Q_1 + \cdots + P_nQ_n$, and in particular it divides A. As $\deg(D) \geq \deg(A)$, this means that $A = cD$ with $c \in K$ nonzero. As both A and D are monic, this implies that $A = D$.

□

Proposition 28.8 *Let $A, B, C \in K[X]$ be such that $gcd(A, B) = gcd(A, C) = 1$. Then A and BC are relatively prime.*

Proof. As A and B are relatively prime, by theorem 28.7, there exist polynomials $U, V \in K[X]$ such that $UA + VB = 1$. As A and C are relatively prime, by theorem 28.7, there exist polynomials $P, Q \in K[X]$ such that $PA + QB = 1$. So we get :

$$1 = (UA+VB)(PA+QC) = UPA^2+UQAC+VPAB+VQBC = (UPA+UQC+VPB)A+(VQ)BC.$$

This proves that the greatest common divisor of A and BC divides 1, hence that it is equal to 1.

□

29 Polynomials of matrices

Notation 29.1 Let $P(X) = \sum_{i=0}^{d} a_i X^i \in K[X]$, and let A be a $n \times n$ matrix. Then the notation $P(A)$ means :

$$a_d A^d + \cdots + a_1 A + a_0 I_n.$$

(That is, $P(A) = \sum_{i=0}^{d} a_i A^i$, where by convention $A^0 = I_n$.)

We have a similar notation for $T \in \mathrm{End}(V)$: We set $P(T) = \sum_{i=0}^{d} a_i T^i$, where by convention $T^0 = \mathrm{id}_V$.

Remark 29.2 We have $(PQ)(T) = P(T) \circ Q(T)$. (Indeed, this is clear if P and Q are monomials, and then this extends to all polynomials by linearity.) As $PQ = QP$, this implies that $P(T) \circ Q(T) = Q(T) \circ P(T)$. In particular, for every $P \in K[X]$, T and $P(T)$ commute (i.e. $T \circ P(T) = P(T) \circ T$).

Lemma 29.3 *Let $T \in \mathrm{End}(V)$, with V a vector space of any dimension, and $P \in K[X]$. If $\lambda \in K$ is an eigenvalue of T and $P(T) = 0$, then $P(\lambda) = 0$.*

Similarly, let $A \in M_n(K)$ and $P \in K[X]$. If $\lambda \in K$ is an eigenvalue of A and $P(A) = 0$, then $P(\lambda) = 0$.

Proof. Let $\vec{v}$ be an eigenvector of T for the eigenvalue λ. Then $T(\vec{v}) = \lambda\vec{v}$, so, for every $i \geq 0$, $T^i(\vec{v}) = \lambda^i \vec{v}$. Hence $0 = P(T)(\vec{v}) = P(\lambda)\vec{v}$. As $\vec{v} \neq 0$ (by definition of an eignevector), this implies that $P(\lambda) = 0$.

□

Remark 29.4 The converse is totally false. For example, if $A = I_n$ and $P = X(X-1)$, then $P(A) = 0$ and $P(0) = 0$, but 0 is not an eigenvalue of A.

Although it looks like a weird technical result, the following theorem is absolutely fundamental in reduction theory.

Theorem 29.5 *Let T be an endomorphism of a vector space V (if any dimension), and let $P \in K[X]$ such that $P(T) = 0$. Suppose that $P = P_1 \dots P_n$, and that, for each $i \not\mid j$, the polynomials P_i and P_j are relatively prime. Then :*

$$V = \mathrm{Ker}(P_1(T)) \oplus \cdots \oplus \mathrm{Ker}(P_n(T)).$$

Example 29.6 In problem 3 of problem set 4, we defined a projection as an endomorphism T of V such that $T^2 = T$. This condition can also be written as $P(T) = 0$, where P is the polynomial $X^2 - X$. As $X^2 - X = X(X-1)$ and the polynomials X and $X-1$ are relatively prime, the theorem implies that $V = \mathrm{Ker}(T) \oplus \mathrm{Ker}(T - \mathrm{id} - V)$. If V is

finite-dimensional, this implies the conclusion of problem 3 of PS4 : if we choose a basis of $\mathrm{Ker}(T - \mathrm{id}_V)$ and a basis of $Ker(T)$ and put them together to get a basis of V, then the matrix of T in this basis is of the form

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix},$$

where $r = \mathrm{rank}(T)$.

Example 29.7 In problem 4 of problem set 4, we defined an involution as an endomorphism T of V such that $T^2 = \mathrm{id}_V$. This condition can also be written as $P(T) = 0$, where P is the polynomial $X^2 - 1$. As $X^2 - X = (X+1)(X-1)$ and the polynomials $X+1$ and $X-1$ are relatively prime, the theorem implies that $V = \mathrm{Ker}(T + \mathrm{id}_V) \oplus \mathrm{Ker}(T - \mathrm{id} - V)$. If V is finite-dimensional, this implies the conclusion of problem 4 of PS4 : if we choose a basis of $\mathrm{Ker}(T - \mathrm{id}_V)$ and a basis of $Ker(T + \mathrm{id}_V)$ and put them together to get a basis of V, then the matrix of T in this basis is of the form

$$\begin{pmatrix} I_r & 0 \\ 0 & -I_{n-r} \end{pmatrix},$$

where $r = \dim \mathrm{Ker}(T - \mathrm{id}_V)$.

Proof. We reason by induction on n. The result is obvious if $n = 1$, because then $P_1 = P$, so $P_1(T) = 0$, so $V = \mathrm{Ker}(P_1(T))$.

Let's prove the theorem for $n = 2$. Write $W_1 = \mathrm{Ker}(P_1(T))$ and $W_2 = \mathrm{Ker}(P_2(T))$. First we note that W_1 and W_2 are stable by T (i.e. $T(W_1) \subset W_1$ and $T(W_2) \subset W_2$). Indeed, this follows from lemma 27.9 and from the fact that T commutes with $P_1(T)$ and $P_2(T)$.

By Bezout's theorem, there exists polynomials $Q_1, Q_2 \in K[X]$ such that $Q_1P_1 + Q_2P_2 = 1$. This gives $Q_1(T) \circ P_1(T) + Q_2(T) \circ P_2(T) = \mathrm{id}_V$. So for every $\vec{v} \in V$, we have

$$\vec{v} = Q_1(T)(\vec{w}_2) + Q_2(T)(\vec{w}_1),$$

with $\vec{w}_2 = P_1(T)(\vec{v})$ and $\vec{w}_2 = P_2(T)(\vec{v})$. As

$$P_2(T)(\vec{w}_2) = P_2(T)(P_1(T)(\vec{v})) = (P_2P_1)(T)(\vec{v}) = P(T)(\vec{v}) = 0,$$

we have $\vec{w}_2 \in W_2$. Similarly, $\vec{w}_1 \in W_1$. We noted above that W_1 and W_2 are both stable by T, so $Q_1(T)(\vec{w}_2) \in W_2$ and $Q_2(T)(\vec{w}_1) \in W_1$. So we have proved that $V = W_1 + W_2$.

To finish the proof in the case $n = 2$, we need to show that $W_1 \cap W_2 = \{0\}$. Let $\vec{v} \in W_1 \cap W_2$. Then $P_1(T)(\vec{v}) = P_2(T)(\vec{v}) = 0$, so

$$\vec{v} = Q_1(T)(P_1(T)(\vec{v})) + Q_2(T)(P_2(T)(\vec{v})) = 0.$$

Now suppose that $n \geq 3$ and that we have proved the theorem for all smaller values of n. Let $Q_2 = P_2 \dots P_n$. Then P_1 and Q_2 are relatively prime (by proposition 28.8), so, by the case $n = 2$ of the theorem, we have :

$$V = \mathrm{Ker}(P_1(T)) \oplus \mathrm{Ker}(Q_2(T)).$$

Let $W = \mathrm{Ker}(Q_2(T))$, and let $U \in \mathrm{End}(W)$ be the restriction of T to W (we know that W is stable by T, because T and $Q(T)$ commute). Then $Q_2(U) = 0$ and $Q_2 = P_2 \dots P_n$, so, by the induction hypothesis :

$$W = \mathrm{Ker}(P_2(U)) \oplus \cdots \oplus \mathrm{Ker}(P_n(U)) = \mathrm{Ker}(P_2(T)) \oplus \cdots \oplus \mathrm{Ker}(P_n(T)).$$

These two equalities gives the result.

□

4/13/2017

Corollary 29.8 *Let $T \in \mathrm{End}(V)$, with V finite-dimensional. Then T is diagonalizable if and only if there exists a polynomial $P \in K[X]$ such that all the roots of P are simple and in K, and such that $P(T) = 0$.*

We have a similar result for matrices : $A \in M_n(K)$ is diagonalizable if and only if there exists a polynomial $P \in K[X]$ such that all the roots of P are simple and in K, and such that $P(Q) = 0$.

Proof. Suppose that T is diagonalizable, and choose a basis $\mathfrak{B} = (\vec{v}_1, \dots, \vec{v}_n)$ of V in which the matrix of T is diagonal. Let $\lambda_1, \dots, \lambda_r$ be the eigenvalues of T (with the λ_i distinct), and let $P(X) = (X - \lambda_1) \dots (X - \lambda_r)$. If $i \in \{1, \dots, n\}$, then there exists j such that $T(\vec{v}_i) = \lambda_j \vec{v}_i$, hence $(T - \lambda_j \mathrm{id}_V)(\vec{v}_i) = 0$, hence $P(T)(\vec{v}_i) = 0$. As this is true for every i and $\mathfrak{B}$ is a basis of V, we have $P(T) = 0$. Also, the polynomial P has all its roots simple and in K by construction.

Conversely, let P be a polnynomial such that all the roots of P are simple and in K, and assume that $P(T) = 0$. By the hypothesis on the roots of P, we have $P(X) = (X - \lambda_1) \dots (X - \lambda_r)$, with the λ_i distinct. This means that, for $i \neq j$, $X - \lambda_i$ and $X - \lambda_j$ are relatively prime. As $P(T) = 0$, theorem 29.5 above gives

$$V = \mathrm{Ker}(T - \lambda_1 \mathrm{id}_V) \oplus \cdots \oplus \mathrm{Ker}(T - \lambda_r \mathrm{id}_V).$$

This implies that T is diagonalizable. (Take bases of each $\mathrm{Ker}(T - \lambda_i \mathrm{id}_V)$, and then take their union to get a basis of V where the matrix of T is diagonal.)

□

Example 29.9 If $A \in M_n(\mathbb{C})$ is such that $A^{100} = I_n$, then A is diagonalizable. Indeed, the polynomial $X^{100} - 1$ has simple roots, and these roots are all in $\mathbb{C}$.

Corollary 29.10 *If $T \in \mathrm{End}(V)$ is diagonalizable and W is a subspace of V such that $T(W) \subset W$, then the endomorphism T_W of W induced by T is diagonalizable.*

Proof. Let $P \in K[X]$ be a polynomial such that all the roots of P are simple and in K and that $P(T) = 0$. Then $P(T_W) = 0$, so by the corollary above T_W is diagonalizable.

□

Now we can prove the result mentioned just after theorem 27.10.

Corollary 29.11 *Let T and U be two endomorphism of a finite-dimensional vector space V. Suppose that T and U are both diagonalizable and that $T \circ U = U \circ T$. Then there exists a basis $\mathfrak{B}$ of V such that both $[T]_{\mathfrak{B}}$ and $[U]_{\mathfrak{B}}$ are diagonal.*

Similarly, if A and B are diagonalizable matrices in $M_n(K)$ and if $AB = BA$, then there exists $S \in M_n(K)$ invertible such that both SAS^{-1} and SBS^{-1} are diagonal.

Proof. Let $\lambda_1, \ldots, \lambda_r$ be the eigenvalues of T, and let $E_{\lambda_i} = \mathrm{Ker}(T - \lambda_i \mathrm{id}_V)$. As T is diagonalizable, we have :

$$V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_r}.$$

By lemma 27.9, each E_{λ_i} is stable by U. By the corollary above, the endomorphism U_i of E_{λ_i} induced by U is diagonalizable. For every i, we choose a basis $\mathfrak{B}_i$ of E_{λ_i} such that the matrix of U_i in $\mathfrak{B}_i$ is diagonal. Let $\mathfrak{B} = \mathfrak{B}_1 \cup \ldots \mathfrak{B}_r$. Then $\mathfrak{B}$ is a basis of V, and the matrices of T and U in $\mathfrak{B}$ are both diagonal.

□

30 Cayley-Hamilton theorem and minimal polynomial

Theorem 30.1 *(Cayley-Hamilton theorem.) Let $T \in \mathrm{End}(V)$ with V finite-dimensional. Then $f_T(T) = 0$.*

We have a similar result for matrices : For every $A \in M_n(K)$, $f_A(A) = 0$.

Remark 30.2 To prove this theorem for $A \in M_n(K)$, it is very tempting to write $f_A(A) = \det(AI_n - A) = \det(0) = 0$. But this doesn't work. To see why, note that $XI_n - A$ is the matrix

$$\begin{pmatrix} X - a_{11} & \ldots & -a_{1n} \\ & \ddots & \\ -a_{n1} & \ldots & X - a_{nn} \end{pmatrix},$$

where $A = (a_{ij})$. So if we replace X by A, we don't get the zero matrix, we get the matrix

$$\begin{pmatrix} A - a_{11}I_n & \ldots & -a_{1n}I_n \\ & \ddots & \\ -a_{n1}I_n & \ldots & A - a_{nn}I_n \end{pmatrix}$$

(with entries in $M_n(K)$!). Also, it is not clear that determinants will work well in that case (we were using commutativity in the proofs), and even if they do it is not clear that the determinant of this matrix is zero.

Proof. As usual, it's enough to prove the result for endomorphisms. We will give two different proofs.

Here is the first proof. Suppose that the characteristic polynomial $f_T(X)$ of T has all its roots in K. [22] Then, by theorem 27.4, we can find a basis $\mathfrak{B} = (\vec{v}_1, \dots, \vec{v}_n)$ such that $[T]_{\mathfrak{B}}$ is upper triangular. Let $\lambda_1, \dots, \lambda_n$ be the diagonal entries of $[T]_{\mathfrak{B}}$. Then we have

$$T(\vec{v}_i) - \lambda_i \vec{v}_i \in \mathrm{Span}(\vec{v}_1, \dots, \vec{v}_{i-1})$$

for every i. Also, by example 25.8,

$$f_T(X) = (X - \lambda_1) \dots (X - \lambda_n).$$

For every $i \in \{1, \dots, n\}$, let $W_i = \mathrm{Span}(\vec{v}_1, \dots, \vec{v}_i)$. We have $T(W_i) \subset W_i$ because the matrix of T in $\mathfrak{B}$ is upper triangular. Also, by the first displayed formula above,

$$(T - \lambda_i \mathrm{id}_V)(W_i) \subset W_{i-1}$$

for every $i \geq 1$, where we set $W_0 = \{0\}$. Hence

$$f_T(T)(W_n) = ((T - \lambda_1 \mathrm{id}_V) \circ \dots \circ (T - \lambda_n \mathrm{id}_V))(W_n) \subset$$

$$\subset ((T - \lambda_1 \mathrm{id}_V) \circ \dots \circ (T - \lambda_{n-1} \mathrm{id}_V))(W_{n-1}) \subset \dots \subset (T - \lambda_1 \mathrm{id}_V)(W_1) = \{0\}$$

As $W_n = V$, this means that $f_T(T) = 0$.

The second proof is a bit longer but doesn't require us to enlarge K. Let $\vec{v} \in V$ be nonzero. We want to prove that $f_T(T)(\vec{v}) = 0$. Let d be the smallest positive integer such that the family $(\vec{v}, T(\vec{v}), \dots, T^{d-1}(\vec{v}))$ is linearly independent, and let $W = \mathrm{Span}(\vec{v}, T(\vec{v}), \dots, T^{d-1}(\vec{v}))$. By (iii) of proposition 7.5, $T^d(\vec{v}) \in W$. Let $a_0, \dots, a_{d-1} \in K$ such that

$$T^d(\vec{v}) = \sum_{i=0}^{d-1} a_i T^i(\vec{v}),$$

and let $f(T) = X^d - \sum_{i=0}^{d-1} a_i X^i \in K[X]$.

First let's prove by induction on r that $T^r(\vec{v}) \in W$ for every $r \geq 0$. We already know it for $r \leq d$, so let's assume that $r > d$ a,d that the result is known for smaller values of r. Then

$$[T^r(\vec{v}) = T^{r-d}(T^d(\vec{v})) = T^{r-d}(\sum_{i=0}^{d-1} a_i T^i(\vec{v})) = \sum_{i=0}^{d-1} a_i T^{r-d+i}(\vec{v}).$$

As every term in the last sum is in W by the induction hypothesis, $T^r(\vec{v})$ is also in W.

By the previous paragraph and the definition of W, we have $T(W) \subset W$. Let T_W be the endomorphism of W induced by T. By proposition 25.9, $f_{T_W}(X)$ divides $f_T(X)$. So if we show that $f_{T_W}(T)(\vec{v})$, it will follow that $f_T(T)(\vec{v}) = 0$ (thanks to remark 29.2). Remember the polynomial $f(T) = X^d - \sum_{i=0}^{d-1} a_i X^i$ from above. By the choice of the a_i, we have $f(T)(\vec{v}) = 0$. So to finish the proof, it suffices to show that $f(X) = f_{T_W}(X)$.

[22] This can always be achieved by formally adding the roots of $f_T(X)$ to K, although making this mathematical construction precise requires things more advanced than MAT 217. It is also always true if $K = \mathbb{C}$.

To prove this last equality, we first note that the family $(\vec{v}, T(\vec{v}), \ldots, T^{d-1}(\vec{v}))$ is a basis of W (it's linearly independent by the choice of d and spans W by definition of W). In this basis, the matrix of T_W is

$$A = \begin{pmatrix} 0 & \ldots & 0 & a_0 \\ 1 & \ddots & \vdots & \vdots \\ 0 & \ddots & 0 & a_{d-2} \\ 0 & 0 & 1 & a_{d-1} \end{pmatrix}.$$

(This kind of matrix is called a *companion matrix*.) So we just need to prove that $f_A(X) = f(X)$. We do this by induction on d. The result is easy for $d = 1$, so let's suppose that $d \geq 2$ and that we know it for $d-1$. We have :

$$f_A(X) = \det \begin{pmatrix} X & \ldots & 0 & -a_0 \\ -1 & \ddots & \vdots & \vdots \\ 0 & \ddots & X & -a_{d-2} \\ 0 & 0 & -1 & X - a_{d-1} \end{pmatrix}.$$

Developing along the first row gives :

$$f_A(X) = X \det \begin{pmatrix} X & 0 & 0 & -a_1 \\ -1 & \ddots & & \vdots \\ 0 & \ddots & X & -a_{d-2} \\ 0 & 0 & -1 & X - a_{d-1} \end{pmatrix} - (-1)^{1+d} a_0 \det \begin{pmatrix} -1 & X & 0 & 0 \\ 0 & -1 & \ddots & 0 \\ \vdots & \ddots & \ddots & X \\ 0 & \ldots & 0 & -1 \end{pmatrix}.$$

Using the induction hypothesis to calculate the first $(d-1) \times (d-1)$ determinant, we get :

$$f_A(X) = X(X^{d-1} - \sum_{i=1}^{d-1} a_i X^{i-1}) - a_0 = f(X).$$

□

Corollary 30.3 *Let $T \in \mathrm{End}(V)$ with V a n-dimensional K-vector space. Then T is nilpotent if and only if $f_T(X) = X^n$. Also, if T is nilpotent, then $T^n = 0$.* [23]

For matrices, this becomes : Let $A \in M_n(K)$. Then A is nilpotent if and only if $f_A(X) = X^n$. Also, if A is nilpotent, then $A^n = 0$.

Proof. Let's prove the result for endomorphisms. If $f_T(X) = X^n$, then $T^n = 0$ by theorem 30.1, and we also get the second sentence.

Now suppose that T is nilpotent. We show by induction on $n := \dim(V)$ that $f_T(X) = X^n$. If $n = 1$, then $T = 0$, so the result is obvious. Assume that $n \geq 2$ and that the

[23] Remember that we say that T is nilpotent if there exists $N \geq 1$ such that $T^N = 0$.

result is know for $n-1$. As T is nilpotent, it cannot be invertible, so $\mathrm{Ker}(T) \neq \{0\}$. Choose a nonzero $\vec{v}_1 \in \mathrm{Ker}(T)$, and complete it to a basis $(\vec{v}_1, \ldots, \vec{v}_n)$ of V. In this basis, the matrix of T is of the form

$$A = \left(\begin{array}{c|c} 0 & * \\ \hline 0 & B \end{array}\right),$$

with $B \in M_{n-1}(K)$. For every N, we have

$$A^N = \left(\begin{array}{c|c} 0 & * \\ \hline 0 & B^N \end{array}\right).$$

As A is nilpotent, B is also nilpotent, so the induction hypothesis gives $f_B(T) = T^{n-1}$. By proposition 25.9, $f_A(X) = Xf_B(X) = X^n$.

□

4/18/2017

Another application of the Cayley-Hamilton theorem is to the minimal polynomial of an endormorphism or matrix. First we must define this polynomial.

Proposition-Definition 30.4 *Let T be an endomorphism of a finite-dimensional vector space V. Then the set I_T of polynomials $f(X) \in K[X]$ such that $f(T) = 0$ is a ideal of $K[X]$ different from $\{0\}$.*

By corollary 28.5, there is a unique monic polynomial $f(X) \in K[X]$ such that $I_T = (f(X))$. This $f(X)$ is called the minimal polynomial *of T.*

We have a similar definition for matrices.

In other words, the minimal polynomial of T is the monic polynomial of smallest degree $f(X)$ such that $f(T) = 0$, and it divides every other polynomial $g(X)$ such that $g(T) = 0$.

Proof. It is a straightforward verification to check that I_T is an ideal of $K[X]$, but it is not totally obvious that it is not equal to $\{0\}$. So we must find a nonzero polynomial $f(X)$ such that $f(T) = 0$. As $\mathrm{End}(V)$ is a finite-dimensional K-vector space and the family $(T^n, n \geq 0)$ of elements of $\mathrm{End}(V)$ is infinite, this family is free. So there exists $N \geq 0$ and scalars $a_0, \ldots, a_N \in K$ not all zero such that $\sum_{r=0}^{N} a_r T^r = 0$. So $f(X) := \sum_{r=0}^{N} a_r X^r$ is a nonzero element of I_T.

□

The following corollary follows immediately from theorem 30.1 (which says, in the notation of the definition above, that the characteristic polynomial $f_T(X)$ is in I_T).

Corollary 30.5 *Let $T \in \mathrm{End}(V)$ be as above, and let $f(X)$ be its minimal polynomial. Then $f(X)$ divides $f_T(X)$. In particular,* $\deg(f(X)) \leq \dim(V)$.

We have a similar statement for matrices.

Example 30.6 The minimal and characteristic polynomials of an endomorphism (or matrix) can be different. For example, if $A = I_n$, the the minimal polynomial of A is $X - 1$, but its characteristic polynomial is $(X-1)^n$. If T is a projection and $T \neq 0, \mathrm{id}_V$, then its minimal polynomial is $X(X-1)$, but its characteristic polynomial is $X^{n-r}(X-1)^r$, where $r = \mathrm{rank}(T)$. If T is an involution such that $T \neq \pm\mathrm{id}_V$, then its minimal polynomial is $(X-1)(X+1)$, but its characteristic polynomial is $(X-1)^r(X+1)^{n-r}$, where $r = \dim(\mathrm{Ker}(T - \mathrm{id}_V))$.

31 Characteristic subspaces

In this section, V is always a finite-dimensional vector space.

Definition 31.1 Let $T \in \mathrm{End}(V)$, and let λ be an eigenvalue of T. We write $f_T(X) = (X-\lambda)^r P(X)$, with $P(\lambda) \neq 0$. Then :

(1) The *characteristic space* of T for the eigenvalue λ is $\mathrm{Ker}((T - \lambda\mathrm{id}_V)^r)$. We sometimes denote it by C_λ.

(2) The *algebraic multiplicity* of λ is $m_a(\lambda) = r$ (i.e. the multiplicity of λ as a root of $f_T(X)$).

(3) The *geometric multiplicity* of λ is $m_g(\lambda) = \dim(\mathrm{Ker}(T - \lambda\mathrm{id}_V))$ (i.e. the dimension of the λ-eigenspace of T).

Proposition 31.2 *Let T, λ and $f_T(X) = (X-\lambda)^r P(X)$ be as in the definition above. Then :*

(i) $\mathrm{Ker}(T - \lambda\mathrm{id}_V) \subset \mathrm{Ker}((T - \lambda\mathrm{id}_V)^r)$ *(the characteristic space contains the eigenspace).*

(ii) $\dim \mathrm{Ker}((T - \lambda\mathrm{id}_V)^r) = r = m_a(\lambda)$.

(iii) $m_a(\lambda) \geq m_g(\lambda)$.

Proof. First, note that (i) follows from the fact that $r \geq 1$ (because λ is a root of $f_T(X)$), and that (iii) follows from (ii). So we just need to prove (ii). Let $C_\lambda = \mathrm{Ker}((T - \lambda\mathrm{id}_V)^r)$ and $W = \mathrm{Ker}(P(T))$. As $P(\lambda) \neq 0$, $X - \lambda$ does not divide P, so $gcd(X - \lambda, P) = 1$. By theorem 29.5, we have

$$V = C_\lambda \oplus W.$$

By lemma 27.9, $T(C_\lambda) \subset C_\lambda$ and $T(W) \subset W$. Let $T_1 \in \mathrm{End}(C_\lambda)$ be the restriction of T, and $T_2 \in \mathrm{End}(W)$ be the restriction of T. By proposition 25.9, $f_T(X) = f_{T_1}(X) f_{T_2}(X)$.

On the one hand, $(T_1 - \lambda\mathrm{id}_{C_\lambda})^r = 0$ (by definition of C_λ), so $T_1 - \lambda\mathrm{id}_{C_\lambda}$ is nilpotent, so, by corollary 30.3, its charactistic polynomial is X^d, where $d = \dim(C_\lambda)$. As

$$f_{T_1}(X) = \det(X\mathrm{id} - T_1) = \det((X-\lambda)\mathrm{id} - (T_1 - \lambda\mathrm{id})) = f_{T_1 - \lambda\mathrm{id}}(X - \lambda),$$

we get $f_{T_1}(X) = (X-\lambda)^d$. As $f_{T_1}(X)$ divides $f_T(X)$, $d \leq r$.

On the other hand, $P(T_2) = 0$ (by definition of W). As $P(\lambda) \neq 0$, this implies by lemma 29.3 that λ is not an eigenvalue of T_2, so that $f_{T_2}(\lambda) \neq 0$. As $f_{T_2}(X)$ divides $f_T(X) = (X-\lambda)^r P(X)$, we get that $f_{T_2}(X)$ divides $P(X)$.

But now we have $f_T(X) = (X-\lambda)^r P(X) = f_{T_1}(X) f_{T_2}(X)$, with $f_{T_1}(X)$ dividing $(X-\lambda)^r$ and $f_{T_2}(X)$ dividing $P(X)$. The only way this can happen is if $f_{T_1}(X) = (X-\lambda)^r$ and $f_{T_2}(X) = P(X)$. In particuliar,

$$\dim(C_\lambda) = \deg(f_{T_1}(X)) = r,$$

which gives (ii).

□

Proposition 31.3 *Let $T \in \mathrm{End}(V)$. For every eigenvalue λ of T, denote by C_λ the corresponding characteristic space. Assume that $f_T(X)$ has all its roots in K, and call these these roots $\lambda_1, \ldots, \lambda_m$. Then :*

$$V = C_{\lambda_1} \oplus \cdots \oplus C_{\lambda_m}.$$

Proof. If $f_T(X)$ has all its roots in K, then we have $f_T(X) = (X-\lambda_1)^{r_1} \ldots (X-\lambda_m)^{r_m}$. By definition, $C_{\lambda_i} = \mathrm{Ker}((T - \lambda_i \mathrm{id}_V)^{r_i})$. By theorem 29.5, $V = C_{\lambda_1} \oplus \cdots \oplus C_{\lambda_m}$.

□

Corollary 31.4 *Let $T \in \mathrm{End}(V)$, and suppose that $f_T(X)$ has all its roots in K. Denote by E_λ and C_λ the eigenspace and characteristic space corresponding to an eigenvalue λ. Then the following conditions are equivalent :*

(i) T is diagonalizable.

(ii) For every eigenvalue λ of T, $E_\lambda = C_\lambda$.

(iii) For every eigenvalue λ of T, $m_g(\lambda) = m_a(\lambda)$.

Proof. We have seen in proposition 31.2 that, for every eigenvalue λ of T, $E_\lambda \subset C_\lambda$, $m_g(\lambda) = \dim(E_\lambda)$ and $m_a(\lambda) = \dim(C_\lambda)$. This shows that (ii) and (iii) are equivalent.

We have seem in proposition 31.3 that $V = C_{\lambda_1} \oplus \cdots \oplus C_{\lambda_m}$, where $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of T. By definition, T is diagonalizable if and only if $V = E_{\lambda_1} \oplus \ldots E_{\lambda_m}$. As $E_{\lambda_i} \subset C_{\lambda_i}$ for every i, this shows that (i) and (ii) are equivalent.

□

32 Jordan-Chevalley decomposition

Theorem 32.1 *(Jordan-Chevalley decomposition) Let $T \in \mathrm{End}(V)$, with $\dim(V) < +\infty$. Suppose that $f_T(X)$ has all its roots in K.*[24]

Then there exist unique endomorphisms T_d, T_n of V such that :

[24] For example, this is automatically true if $K = \mathbb{C}$.

(a) $T = T_d + T_n$.

(b) $T_d \circ T_n = T_n \circ T_d$.

(c) T_d *is diagonalizable and* T_n *is nilpotent.*

We have a similary theorem for matrices :

Theorem 32.2 *Let* $A \in M_n(K)$. *Suppose that* $f_A(X)$ *has all its roots in* K.
Then there exists unique matrices $A_d, A_n \in M_n(K)$ *such that :*

(a) $A = A_d + A_n$.

(b) $A_d A_n = A_n A_d$.

(c) A_d *is diagonalizable and* A_n *is nilpotent.*

Example 32.3 Let $A = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$. Then A is diagonalizable, so its Jordan-Chevalley decomposition is given by $A = A_d + A_n$, with $A_d = A$ and $A_n = 0$.

Proof. We prove the theorem for endomorphisms. Let $\lambda_1, \dots, \lambda_m$ be the eigenvalues of T, and let $C_{\lambda_1}, \dots, C_{\lambda_m}$ be the corresponding characteristic spaces. Then $V = C_{\lambda_1} \oplus \dots \oplus C_{\lambda_m}$ by proposition 31.3.

Existence : We define T_d and T_n in the following way : for every $i \in \{1, \dots, m\}$, for every $\vec{v} \in C_{\lambda_i}$, $T_d(\vec{v}) = \lambda_i \vec{v}$ and $T_n(\vec{v}) = T(\vec{v}) - \lambda_i \vec{v}$. As $V = C_{\lambda_1} \oplus \dots \oplus C_{\lambda_m}$, this formula defines unique endomorphisms T_d and T_n of V, and we have $T = T_d + T_n$ because this is true on each C_{λ_i}.

Let's check that T_d is diagonalizable. For every i, $C_{\lambda_i} \subset \mathrm{Ker}(T_d - \lambda_i \mathrm{id}_V)$. So

$$V = \sum_{i=1}^{m} \mathrm{Ker}(T_d - \lambda_i \mathrm{id}_V),$$

and so we can find a basis of V made up of eigenvectors of T_d.

Let's check that T_n is nilpotent. Write $f_T(X) = (X - \lambda_1)^{r_1} \dots (X - \lambda_m)^{r_m}$. Then $(T - \lambda_i \mathrm{id}_V)^{r_i}(C_{\lambda_i}) = 0$ for every i (by definition of C_i), so $T_n^{r_i}(C_{\lambda_i}) = 0$ for every i. As $V = C_{\lambda_1} \oplus \dots \oplus C_{\lambda_m}$, we see that, if $N \geq \sup(r_1, \dots, r_m)$, then $T_n^N = 0$.

Finally, let's check that $T_d \circ T_n = T_n \circ T_d$. This is true on each C_{λ_i} (because T_d is just $\lambda_i \mathrm{id}_{C_{\lambda_i}}$ on C_{λ_i}), so it's true on all of V because $V = C_{\lambda_1} \oplus \dots \oplus C_{\lambda_m}$.

4/20/2017

Uniqueness : Suppose that we have another couple (U_d, U_n) satisfying conditions (a), (b) and (c). By (b), U_d commutes with U_n, so by (a) it also commutes with $T = U_d + U_n$,

and hence, by lemma 27.9, $U_d(C_{\lambda_i}) \subset C_{\lambda_i}$ for every i. As T_d is just $\lambda_i \mathrm{id}_{C_{\lambda_i}}$ on C_{λ_i}, we see that $T_d \circ U_d = U_d \circ T_d$ on C_{λ_i}. As $V = C_{\lambda_1} \oplus \cdots \oplus C_{\lambda_m}$, this implies that $T_d \circ U_d = U_d \circ T_d$.

By corollary 29.11, there exists a basis $\mathfrak{B}$ of V such that both $[T_d]_{\mathfrak{B}}$ and $[U_d]_{\mathfrak{B}}$ are diagonal. But then $[U_d - T_d]_{\mathfrak{B}} = [U_d]_{\mathfrak{B}} - [T_d]_{\mathfrak{B}}$ is also diagonal, and so $U_d - T_d$ is diagonalizable.

Now we know that U_d commutes with T and T_d, so it also commutes with $T_n = T - T_d$. And T_n commutes with T and U_d, so it also commutes with $U_n = T - U_d$. This implies that $T_n - U_n$ is nilpotent. Indeed, choose N big enough so that $T_n^N = 0$ and $U_n^N = 0$. Then :

$$(T_n - U_n)^{2N} = \sum_{i=0}^{2N} (-1)^{2N-i} \binom{2N}{i} T_n^i \circ U_n^{2N-i}$$

(using the fact that T_n and U_n commute). For every $i \in \{0, \ldots, 2N\}$, either $i \geq N$ and then $T_n^i = 0$, or $2N - i \geq N$ and then $U_n^{2N-i} = 0$. So $(T_n - U_n)^{2N} = 0$.

In summary, we have proved that $U_d - T_d$ is diagonalizable and that $T_n - U_n$ is nilpotent. We also have $T = T_d + T_n = U_d + U_n$, so $T_n - U_n = U_d - T_d$. So $T_n - U_n$ is both diagonalizable and nilpotent. As it is nilpotent, its only eigenvalue is 0, and as it is diagonalizable, its matrix in some basis must therefore be the zero matrix. But then $T_n - U_n$ has to be 0, so $T_n = U_n$, and then $U_d = T_d$.

□

33 Jordan normal/canonical form

We now push the analysis of the previous section one step further and show that, if all the roots of $f_T(X)$ are in K, then we can find a basis of V where the matrix of T is very simple. The base case is that of nilpotent endomorphisms.

Theorem 33.1 *Let V be a n-dimensional vector space, and let $T \in \mathrm{End}(V)$ be nilpotent. Then we can write $V = V_1 \oplus \cdots \oplus V_r$, with :*

(a) $T(V_i) \subset V_i$ for every $i \in \{1, \ldots, r\}$.

(b) For every $i \in \{1, \ldots, r\}$, there exists a basis $\mathfrak{B}_i$ of V_i such that, if $T_i \in \mathrm{End}(V_i)$ is the restriction of T :

$$[T_i]_{\mathfrak{B}_i} = \begin{pmatrix} 0 & 1 & \ldots & 0 \\ & \ddots & \ddots & \vdots \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}.$$

By taking $\mathfrak{B} = \mathfrak{B}_1 \cup \cdots \cup \mathfrak{B}_r$, we get a basis of V such that

$$[T]_{\mathfrak{B}} = \begin{pmatrix} 0 & c_1 & \ldots & 0 \\ & \ddots & \ddots & \vdots \\ & & \ddots & c_{n-1} \\ 0 & & & 0 \end{pmatrix},$$

where each c_j is either 0 or 1.

Proof. We prove the result by induction on n. If $n = 1$, then $T = 0$ and we are done. So suppose that $n \geq 2$ and that we know the result for smaller values of n.

By corollary 30.3, $f_T(X) = X^n$. Let $f_{min}(X)$ be the minimal polynomial of T. Then $f_{min}(X)$ divides $f_T(X)$, so $f_{min}(X) = X^d$, with $d \leq n$. By definition of the minimal polynomial, $T^{d-1} \neq 0$ (otherwise $f_{min}(X)$ would have to divide X^{d-1}). Choose $\vec{v} \in V$ such that $T^{d-1}(\vec{v}) \neq 0$, and let $W = \text{Span}(\vec{v}, T(\vec{v}), \ldots, T^{d-1}(\vec{v}))$. As $T^d(\vec{v}) = 0 \in W$, we have $T(W) \subset W$. I claim that $(\vec{v}, T(\vec{v}), \ldots, T^{d-1}(\vec{v}))$ is a basis of W. As this family spans W, we just need to show that is is linearly independent. This was done in a problem set and the midterm : Let $a_0, \ldots, a_{d-1} \in K$ be such that $a_0\vec{v} + a_1 T(\vec{v}) + \cdots + a_{d-1}T^{d-1}(\vec{v}) = 0$, and suppose that $a_0, \ldots, a_{d-1}$ are not all 0. Let i be the smaller integer such that $a_i \neq 0$. Then

$$0 = T^{d-1-i}(a_i T^i(\vec{v}) + \cdots + a_{d-1}T^{d-1}(\vec{v})) = a_i T^{d-1}(\vec{v})$$

(the other terms disappear because $T^j(\vec{v}) = 0$ if $j \geq d$), hence $T^{d-1}(\vec{v}) = 0$, contradiction.

Let $U \in \text{End}(W)$ be the restriction of T. In the basis $(T^{d-1}(\vec{v}), \ldots, T(\vec{v}), \vec{v})$ of W, the matrix of U is

$$\begin{pmatrix} 0 & 1 & \ldots & 0 \\ & \ddots & \ddots & \vdots \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}.$$

So W can be our V_1 in the theorem. To invoke the induction hypothesis, we need to find another subspace E of V such that $V = W \oplus E$ and $T(E) \subset E$. We have to be a bit careful in the choice of E if we want the second condition to be true.

Write $(\vec{e}_1, \ldots, \vec{e}_d) = (\vec{v}, T(\vec{v}), \ldots, T^{d-1}(\vec{v}))$, and complete this to a basis $(\vec{e}_1, \ldots, \vec{e}_n)$ of V. Let $(\vec{e}_1^*, \ldots, \vec{e}_n^*)$ be the dual basis, and let

$$E = \{\vec{v} \in V | \forall i \geq 0,\ \vec{e}_d^*(T^i(\vec{v})) = 0\} = \bigcap_{i \geq 0} \text{Ker}(\vec{e}_d^* \circ T^i).$$

This is a subspace of V, and we have to show that it works.

First, if $\vec{v} \in E$, then for every $i \geq 0$,

$$\vec{e}_d^*(T^i(T(\vec{v}))) = \vec{e}_d^*(T^{i+1}(\vec{v})) = 0,$$

so $T(\vec{v}) \in E$. So $T(E) \subset E$.

Second, let's show that $W \cap E = \{0\}$. Let $\vec{v} \in W \cap E$. If $\vec{v} \neq 0$, we can write $\vec{v} = a_1\vec{e}_1 + \cdots + a_r\vec{e}_r$, with $r \leq d$ and $a_r \neq 0$. Then

$$T^{d-r}(\vec{v}) = a_1\vec{e}_{d-r+1} + \cdots + a_r\vec{e}_d,$$

so

$$\vec{e}_d^*(T^{d-r}(\vec{v})) = a_r \neq 0,$$

which contradicts the fact that $\vec{v} \in E$.

It remains to show that $W + E = V$. By corollary 10.6, we have

$$\dim(V) \geq \dim(W + E) = \dim(W) + \dim(E) - \dim(W \cap E) = \dim(W) + \dim(E),$$

so we just need to show that $\dim(E) \geq n - d$. Noting that $T^r = 0$ for $r \geq d$ (because the minimal polynomial of T is X^d), we see that

$$E = \mathrm{Ker}(\vec{e_d^*}) \cap \mathrm{Ker}(e_d^* \circ T) \cap \cdots \cap \mathrm{Ker}(e_d^* \circ T^{d-1}).$$

For every $i \geq 0$, $\vec{e_d^*} \circ T^i$ is a linear transformation from V to K, so its rank is $\leq \dim(K) = 1$, so, by the rank-nullity theorem, $\dim(\mathrm{Ker}(e_d^* \circ T^i)) \geq n - 1$. Note also that, if V_1 and V_2 are subspaces of V, then by corollary 10.6 again,

$$\dim(V_1 \cap V_2) = \dim(V_1) + \dim(V_2) - \dim(V).$$

So finally :

$$\begin{aligned}
\dim(E) &= \dim(\bigcap_{i=0}^{d-1} \mathrm{Ker}(e_d^* \circ T^i)) \\
&= \dim(\mathrm{Ker}(e_d^* \circ T^{d-1})) - \dim(V) + \dim(\bigcap_{i=0}^{d-2} \mathrm{Ker}(e_d^* \circ T^i)) \\
&\geq -1 + \dim(\bigcap_{i=0}^{d-2} \mathrm{Ker}(e_d^* \circ T^i)) \\
&= -1 + \dim(\mathrm{Ker}(e_d^* \circ T^{d-2})) - \dim(V) + \dim(\bigcap_{i=0}^{d-3} \mathrm{Ker}(e_d^* \circ T^i)) \\
&\geq -2 + \dim(\bigcap_{i=0}^{d-3} \mathrm{Ker}(e_d^* \circ T^i)) \\
&= \dots \\
&\geq -(d-1) + \dim(\mathrm{Ker}(e_d^*)) \geq n - d.
\end{aligned}$$

So we've found a subspace E of V such that $V = W \oplus E$ and $T(E) \subset E$. Applying the induction hypothesis to E finishes the proof.

□

Corollary 33.2 *(Jordan normal form) Let $T \in \mathrm{End}(V)$, with V finite-dimensional. Assume that $f_T(X)$ has all its roots in K, and write $f_T(X) = (X - \lambda_1)^{r_1} \dots (X - \lambda_m)^{r_m}$. Then there exists a basis $\mathfrak{B}$ of V such that*

$$[T]_{\mathfrak{B}} = \begin{pmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_m \end{pmatrix},$$

where, for every $i \in \{1, \dots, m\}$, $A_i \in M_{r_i}(K)$ is of the form

$$A_i = \begin{pmatrix} \lambda_i & c_{i,1} & \dots & 0 \\ & \ddots & \ddots & \vdots \\ & & \ddots & c_{i,r_i-1} \\ 0 & & & \lambda_i \end{pmatrix},$$

with the $c_{i,j}$ in $\{0, 1\}$.

Proof. By theorem 29.5, we have $V = V_1 \oplus \dots \oplus V_m$, where $V_i = \mathrm{Ker}((T - \lambda_i \mathrm{id}_V)^{r_i})$.

Fix $i \in \{1, \dots, m\}$. Then $T - \lambda_i \mathrm{id}_V$ sends V_i to itself and is nilpotent on V_i, so, by theorem 33.1, there exists a basis $\mathfrak{B}_i$ of V_i such that the matrix of the endomorphism of V_i given by $T - \lambda_i \mathrm{id}_V$ in $\mathfrak{B}_i$ is

$$\begin{pmatrix} 0 & c_{i,1} & \dots & 0 \\ & \ddots & \ddots & \vdots \\ & & \ddots & c_{i,r_i-1} \\ 0 & & & 0 \end{pmatrix},$$

with the $c_{i,j}$ in $\{0, 1\}$. So $T(V_i) \subset V_i$, and the matrix in $\mathfrak{B}_i$ of the endomorphism of V_i induced by T is

$$A_i := \begin{pmatrix} \lambda_i & c_{i,1} & \dots & 0 \\ & \ddots & \ddots & \vdots \\ & & \ddots & c_{i,r_i-1} \\ 0 & & & \lambda_i \end{pmatrix}.$$

Let $\mathfrak{B} = \mathfrak{B}_1 \cup \dots \cup \mathfrak{B}_m$. Then $\mathfrak{B}$ is a basis of T, and

$$[T]_{\mathfrak{B}} = \begin{pmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_m \end{pmatrix}.$$

□

4/25/2017

34 Matrix of a bilinear form

Remember the following definition :

Definition 34.1 Let V be a K-vector space. A *bilinear form* on V is a function $f : V \times V \to K$ such that f is linear in each variable.

We can use matrices to represent bilinear forms in the following way.

Definition 34.2 Let V be a finite-dimensional K-vector space, let $\mathfrak{B} = (\vec{v}_1, \ldots, \vec{v}_n)$, and let f be a bilinear form of V.

The *matrix* of f in the basis $\mathfrak{B}$ is the matrix $B = (b_{i,j}) \in M_n(K)$ defined by :

$$b_{ij} = f(\vec{v}_i, \vec{v}_j).$$

This matrix has the following property.

Lemma 34.3 *With the notation as in the definition above, we have : For every $\vec{v}, \vec{w} \in V$,*

$$f(\vec{v}, \vec{w}) = [\vec{v}]_{\mathfrak{B}}^T B [\vec{w}].$$

Proof. Writen $[\vec{v}]_{\mathfrak{B}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and $[\vec{w}]_{\mathfrak{B}} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$. Then

$$\vec{v} = x_1\vec{v}_1 + \cdots + x_n\vec{v}_n$$

and

$$\vec{w} = y_1\vec{v}_1 + \cdots + y_n\vec{v}_n,$$

so

$$f(\vec{v}, \vec{w}) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j f(\vec{v}_i, \vec{v}_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j b_{ij}.$$

This also happens to be the unique entry of the 1×1 matrix

$$\begin{pmatrix} x_1 & \ldots & x_n \end{pmatrix} B \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

□

Remark 34.4 If A and B are two matrices in $M_n(K)$ such that, for every $X, Y \in M_{n1}(K)$, $X^T A Y = X^T B Y$, then $A = B$. Indeed, taking X to be the ith column of I_n and Y to be its jth column, we get that the(i, j)-entries of A and B are equal.

In particular, the matrix of f in $\mathfrak{B}$ is the only $n \times n$ matrix that satisfies the conclusion of lemma 34.3.

We also have a change of basis formula (note that it's different from the change of basis formula for the matrix of an endomorphism).

Proposition 34.5 *Let V, $\mathfrak{B}$, f and B be as in defintiion 34.2. Let $\mathfrak{A} = (\vec{w}_1, \ldots, \vec{w}_1)$ be another basis of V, and let A be the matrix of f in this basis. Let P be the matrix with columns $[\vec{w}_1]_{\mathfrak{B}}, \ldots, [\vec{w}_n]_{\mathfrak{B}}$.*

Then

$$A = P^T B P.$$

Proof. Let $\vec{x}, \vec{y} \in V$. By proposition 12.1, we have

$$[\vec{x}]_{\mathfrak{A}}^T (P^T B P)[\vec{y}]_{\mathfrak{A}} = (P[\vec{x}]_{\mathfrak{A}})^T B (P[\vec{y}]_{\mathfrak{A}}) = [\vec{x}]_{\mathfrak{B}} B [\vec{y}]_{\mathfrak{B}} = f(\vec{x}, \vec{y}).$$

□

35 Symmetric bilinear form vs quadratic forms

Definition 35.1 (i) A bilinear form f on a K-vector space V is called *symmetric* if, for every $\vec{v}, \vec{w} \in V$,

$$f(\vec{v}, \vec{w}) = f(\vec{w}, \vec{v}).$$

(ii) A bilinear form f on a K-vector space V is called *antisymmetric* if, for every $\vec{v}, \vec{w} \in V$,

$$f(\vec{v}, \vec{w}) = -f(\vec{w}, \vec{v}).$$

(iii) A matrix $B \in M_n(K)$ is called *symmetric* if $B = B^T$, and *antisymmetric* if $B = -B^T$.

Lemma 35.2 *Suppose that V is finite-dimensional, let $\mathfrak{B}$ be a basis of V, let f be a bilinear form on V and let B be its matrix in $\mathfrak{B}$.*

Then f is symmetric if and only if B is symmetric, and f is antisymmetric if and only B is antisymmetric.

Proof. Let $\vec{v}, \vec{w} \in V$. Then :

$$f(\vec{w}, \vec{v}) = \vec{w}^T B \vec{v} = (\vec{v}^T B^T \vec{w})^T = \vec{v}^T B^T \vec{w}$$

(because ^{T} doesn't change 1×1 matrices). This (and remark 34.4) gives the result.

□

Definition 35.3 Let V be a K-vector space. A *quadratic form* on V is a function $q : V \to K$ such that there exists a symmetric bilinear form f on V satisfying :

$$q(\vec{v}) = f(\vec{v}, \vec{v}),$$

for every $\vec{v} \in V$.

Note that we then have $q(\lambda \vec{v}) = \lambda^2 q(\vec{v})$, for $\vec{v} \in V$ and $\lambda \in K$.

We will now show that, if $\mathrm{char}(K) \neq 2$, the form f in the definition is unique.

Proposition 35.4 *Suppose that* $\mathrm{char}(K) \neq 2$ *(that is, that* 2 *is invertible in* K*), and let* q *be a quadratic form on* V*. Then there exists a unique symmetric bilinear form* f *on* V *such that* $q(\vec{v}) = f(\vec{v}, \vec{v})$*.*

Moreover, we have :

$$f(\vec{v}, \vec{w}) = \frac{1}{2}(q(\vec{v} + \vec{w}) - q(\vec{v}) - q(\vec{w})) = \frac{1}{4}(q(\vec{v} + \vec{w}) - q(\vec{v} - \vec{w}))$$

The symmetric bilinear form f is sometimes called the *polar form* of q.

So, if 2 is invertible in K (for example if $K = \mathbb{R}$ or $\mathbb{C}$), it's the same to give a symmetric bilinear form on V or a quadratic form on V. We'll sometimes speak about the matrix of a quadratic form (in a given basis) for the matrix of the corresponding symmetric bilinear form (in the same basis).

Proof. Existence follows from the definition of quadratic forms.

Let's show uniquemess. Let f be a symmetric bilinear form on V such that $q(\vec{v}) = f(\vec{v}, \vec{v})$. Then :

$$q(\vec{v}+\vec{w}) = f(\vec{v}+\vec{w}, \vec{v}+\vec{w}) = f(\vec{v}, \vec{v})+f(\vec{w}, \vec{w})+f(\vec{v}, \vec{w})+f(\vec{w}, \vec{v}) = q(\vec{v})+q(\vec{w})+2f(\vec{v}, \vec{w}),$$

hence

$$f(\vec{v}, \vec{w}) = \frac{1}{2}(q(\vec{v} + \vec{w}) - q(\vec{v}) - q(\vec{w})).$$

The second formula for f can be checked in the same way, or deduced from the first. □

36 Non-degenerate and definite forms

Definition 36.1 Let f be a symmetric bilinear form on a K-vector spacce V. The *kernel* of f is defined by

$$\mathrm{Ker}(f) = \{\vec{v} \in V | \forall \vec{w} \in V,\ f(\vec{v}, \vec{w}) = 0\}.$$

We say that f is *non-degenerate* if $\mathrm{Ker}(f) = \{0\}$.

Proposition 36.2 *With the notation of the definition above, we have :*

(i) $\mathrm{Ker}(f)$ *is a subspace of* V*, and we have*

$$\mathrm{Ker}(f) = \{\vec{w} \in V | \forall \vec{v} \in V,\ f(\vec{v}, \vec{w}) = 0\}.$$

(ii) If V *is finite-dimensional,* $\mathfrak{B}$ *is a basis of* V *and* B *is the matrix of* f *in* $\mathfrak{B}$*, then :*

$$\vec{v} \in \mathrm{Ker}(f) \Leftrightarrow [\vec{v}]_{\mathfrak{B}} \in \mathrm{Ker}(B).$$

(iii) Consider the map $u_f : V \to V^*$ *sending* $\vec{v} \in V$ *to the linear form* $u_f(\vec{v}) : \vec{w} \mapsto f(\vec{v}, \vec{w})$*. Then* u_f *is linear, and* $\mathrm{Ker}(f) = \mathrm{Ker}(u_f)$*. In particular,* f *is non-degenerate if and only if* u_f *is injective, and if* V *is finite-dimensional, then* f *is non-degenerate if and only if* u_f *is an isomorphism.*

Proof.

(i) The formula for $\mathrm{Ker}(f)$ follows from the fact the f is symmetric, and the fact that $\mathrm{Ker}(f)$ is a subspace follows from (iii). (But it can also be checked directly.)

(ii) Let $\vec{v} \in V$. Then $\vec{v} \in \mathrm{Ker}(f)$ if and only if $X^T(B\vec{v}) = 0$ for every $X \in M_{n1}(K)$. So $B\vec{v} = 0$ implies that $\vec{v} \in \mathrm{Ker}(f)$. By taking for X the columns of the identity matrix I_n, we see that $\vec{v} \in \mathrm{Ker}(f)$ implies that $B\vec{v} = 0$.

(iii) The fact that $u_f(\vec{v})$ (for a given $\vec{v}$) is in V^* (i.e. linear) follows from the linearity of f in the first variable. The fact that u_f is linear (i.e. that $u_f(\lambda\vec{v}_1 + \vec{v}_2) = \lambda u_f(\vec{v}_1) + u_f(\vec{v}_2)$) follows from the linearity of f in the second variable.

Finally, we have :

$$\vec{v} \in \mathrm{Ker}(f) \Leftrightarrow \forall \vec{w} \in W,\ f(\vec{v}, \vec{w}) = 0 \Leftrightarrow \forall \vec{w} \in W,\ (u_f(\vec{v}))(\vec{w}) = 0 \Leftrightarrow u_f(\vec{v}) = 0.$$

□

Definition 36.3 The *isotropic cone* of a symmetric bilinear form is

$$C_f = \{\vec{v} \in V | f(\vec{v}, \vec{v}) = 0\}.$$

We say that f is *definite* if $C_f = \{0\}$.

Remark 36.4 Note that C_f is not a subspace of V in general, but it is a cone, which means that $0 \in C_f$ and $\lambda\vec{v} \in C_f$ if $\vec{v} \in C_f$ and $\lambda \in K$.

Remark 36.5 We have $\mathrm{Ker}(f) \subset C_f$ by the definitions, so a definite form is non-degenerate, but the converse is false. For example, the form

$$f((x_1, x_2), (y_1, y_2)) = x_1y_1 - x_2y_2$$

on K^2 is non-degenerate, but it is not definite.

4/27/2017

37 Orthogonals

We fix a K-vector space V and a symmetric bilinear form f on V.

Definition 37.1 We say that $\vec{v}$ and $\vec{w}$ in V are *orthogonal* (with respect to f) if $f(\vec{v}, \vec{w})$. In that case, we write $\vec{v} \perp \vec{w}$.

If A is a subset of V, we write

$$A^\perp = \{\vec{v} \in V | \forall \vec{w} \in A,\ \vec{v} \perp \vec{w}\}.$$

If $A = \{\vec{v}\}$, we also write $\vec{v}^\perp$ for $A^\perp$.

We say that two subsets A and B of V are *orthogonal to each other* (and we write $A \perp B$) if $B \subset A^\perp$ (i.e. every element of A is orthogonal to every element of B).

Remark 37.2 We have

$$A^\perp = \bigcap_{\vec{v}\in A} \vec{v}^\perp,$$

and $\mathrm{Ker}(f) = V^\perp$.

Lemma 37.3 *(Exercise.) Let A be a subset of V. Then :*

(i) We have

$$A^\perp = \mathrm{Span}(A)^\perp,$$

and this is a subspace of V.

(ii) We have $A \subset (A^\perp)^\perp$.

(iii) If $B \supset A$, then $B^\perp \subset A^\perp$.

Proposition 37.4 *Suppose that V is finite-dimensional. Then, for every subspace W of V :*

(i) $\dim(W) + \dim(W^\perp) = \dim(V) + \dim(W \cap \mathrm{Ker}(f))$.

(ii) $(W^\perp)^\perp = W + \mathrm{Ker}(f)$.

In particular, if f is non-degenerate, then $\dim(W) + \dim(W^\perp) = \dim(V)$ *and* $W = (W^\perp)^\perp$.

Proof.

(i) Remember the linear transformation $u_f : V \to V^*$ sending $\vec{v} \in V$ to the linear form $\vec{w} \mapsto f(\vec{v}, \vec{w})$. Let $T = u_{f|W} : W \to V^*$. We've seen in the remark above that $\mathrm{Ker}(u_f) = \mathrm{Ker}(f)$, so $\mathrm{Ker}(T) = W \cap \mathrm{Ker}(f)$. The rank-nullity theorem gives :

$$\dim(W) = \dim(W \cap \mathrm{Ker}(T)) + \dim(\mathrm{Im}(T)).$$

Note that

$$W^\perp = \{\vec{v} \in V | \forall \vec{w} \in W,\ T(\vec{w})(\vec{v}) = 0\} = \bigcap_{\vec{w}\in W} \mathrm{Ker}(T(\vec{w})) = \bigcap_{\varphi\in \mathrm{Im}(T)} \mathrm{Ker}(\varphi).$$

Let $(\vec{v}_1^*, \dots, \vec{v}_r^*)$ be a basis of $\mathrm{Im}(T)$, and complete it to a basis $(\vec{v}_1^*, \dots, \vec{v}_n^*)$ of V^*. Let $(\vec{v}_1, \dots, \vec{v}_n)$ be the basis of V such that $(\vec{v}_1^*, \dots, \vec{v}_n^*)$ is its dual basis.[25] I claim that $(\vec{v}_{r+1}, \dots, \vec{v}_n)$ is a basis of $W^\perp$. All these vectors are in $W^\perp$ because

$$W^\perp = \bigcap_{\varphi\in \mathrm{Im}(T)} \mathrm{Ker}(\varphi) = \bigcap_{i=1}^{r} \mathrm{Ker}(\vec{e}_i^*),$$

[25] This exists for the following reason : We have an isomorphism $V \to V^{**}$ given by $\vec{x} \mapsto (\varphi \mapsto \varphi(\vec{x}))$, and $(\vec{v}_1, \dots, \vec{v}_n)$ is the inverse image by this isomorphism of the dual basis of $(\vec{v}_1^*, \dots, \vec{v}_n^*)$.

so we just need to show that they span $W^\perp$. Let $\vec{v} \in W^\perp$, and write $\vec{v} = a_1\vec{v}_1 + \cdots + a_n\vec{v}_n$. Then if $1 \leq i \leq r$, we have $a_i = e_i^*(\vec{v}) = 0$. So $\vec{v} = a_{r+1}\vec{v_r^*} + \cdots + a_n\vec{v}_n \in \mathrm{Span}(\vec{v}_{r+1}, \ldots, \vec{v}_n)$.

Finally, we have proved that $\dim(W^\perp) = \dim(V) - \dim(\mathrm{Im}(T))$, so we get :

$$\dim(W) = \dim(W \cap \mathrm{Ker}(f)) + \dim(V) - \dim(W^\perp),$$

which is what we wanted.

(ii) We already know that $W \subset (W^\perp)^\perp$ and $Ker(f) = V^\perp \subset (W^\perp)^\perp$, so $W + \mathrm{Ker}(f) \subset (W^\perp)^\perp$. Also, by (i) (for $W^\perp$),

$$\dim((W^\perp)^\perp) = \dim(V) + \dim(W^\perp \cap \mathrm{Ker}(f)) - \dim(W^\perp).$$

As $\mathrm{Ker}(f) = V^\perp \subset W^\perp$, this simplifies to

$$\dim((W^\perp)^\perp) = \dim(V) - \dim(W^\perp) + \dim(\mathrm{Ker}(f)).$$

Using (i) again (this time for W) gives

$$\dim((W^\perp)^\perp) = \dim(V) - (\dim(V) - \dim(W) + \dim(W \cap \mathrm{Ker}(f))) + \dim(\mathrm{Ker}(f))$$

$$= \dim(W) + \dim(\mathrm{Ker}(f)) - \dim(W \cap \mathrm{Ker}(f)) = \dim(W + \mathrm{Ker}(f)),$$

and this implies the result.

□

Corollary 37.5 *Suppose that V is finite-dimensional and that f is* definite*, and let W be a subspace of W. Then $W \cap W^\perp = \{0\}$, so we have :*

$$W \oplus W^\perp = V.$$

38 Orthogonal bases

We fix a K-vector space V and a symmetric bilinear form f on V. "Orthogonal" means "orthogonal with respect to f".

Definition 38.1 We say that a fanily $(\vec{v}_i)_{i \in I}$ of vectors of $\vec{v}$ is *orthogonal* if $\vec{v}_i \perp \vec{v}_j$ for $i \neq j$. If this family is a basis of V, we call it an *orthogonal basis.*

Remark 38.2 If $(\vec{v}_1, \ldots, \vec{v}_r)$ is an orthogonal family and $\vec{v} = a_1\vec{v}_1 + \cdots + a_r\vec{v}_r$, $\vec{w} = b_1\vec{v}_1 + \cdots + b_r\vec{v}_r$, then

$$f(\vec{v}, \vec{w}) = \sum_{i=1}^{r}\sum_{j=1}^{r} a_i b_j f(\vec{v}_i, \vec{v}_j) = \sum_{i=1}^{r} a_i b_i f(\vec{v}_i, \vec{v}_i).$$

Also, by definition, the matrix of f in an orthogonal basis (if V is finite-dimensional) is a diagonal matrix.

Lemma 38.3 *If f is* definite*, then every orthogonal family made up of nonzero vectors is linearly independent.*

This is not true in general. For example, if we have $\vec{v} \neq 0$ such that $f(\vec{v}, \vec{v}) = 0$, then the family $(\vec{v}, \vec{v})$ is orthogonal but not linearly independent.

Proof. Suppose that we have $i_1, \ldots, i_r \in I$ distinct and $a_1, \ldots, a_r \in K$ such that $a_1\vec{v}_{i_1} + \cdots + a_r\vec{v}_{i_r} = 0$. Let $s \in \{1, \ldots, r\}$. Then :

$$0 = f(\vec{v}_{i_s}, a_1\vec{v}_{i_1} + \cdots + a_r\vec{v}_{i_r}) = a_s f(\vec{v}_{i_s}, \vec{v}_{i_s}).$$

By assumption, $\vec{v}_{i_s} \neq 0$, so $f(\vec{v}_{i_s}, \vec{v}_{i_s}) \neq 0$ as f is definite. Hence $a_s = 0$.

□

Proposition 38.4 *Suppose that V is finite-dimensional and that* $\mathrm{char}(K) \neq 2$*. Then it has an orthogonal basis.*

In matrix terms, this says : If $B \in M_n(K)$ is a symetric matrix, then there exists an invertible matrix $P \in M_n(K)$ and a diagonal matrix $D \in M_n(K)$ such that $B = P^T DP$. (Use the change of basis formula for the matrix of f and remark 38.2.)

Proof. By induction on $\dim(V)$. If $\dim(V) = 1$, any basis is orthogonal. So assume that $\dim(V) \geq 2$ and that the result is known in smaller dimensions. If $f(\vec{v}, \vec{v}) = 0$ for every $\vec{v} \in V$, then $f = 0$ by proposition 35.4, so every basis is orthogonal and we are done. Otherwise, choose $\vec{v} \in V$ such that $f(\vec{v}, \vec{v}) \neq 0$. Let $W = \vec{v}^T = (\mathrm{Span}(\vec{v}))^T$. Then we have $\mathrm{Span}(\vec{v}) \cap W = \mathrm{Span}(\vec{v}) \cap \mathrm{Ker}(f) = \{0\}$, so by proposition 37.4 $\dim(W) = \dim(V) - 1$, and so $\dim(\mathrm{Span}(\vec{v}) + W) = \dim(V)$, which implies that $V = \mathrm{Span}(\vec{v}) \oplus W$. By the induction hypothesis, we can find an orthogonal basis $(\vec{v}_2, \ldots, \vec{v}_n)$ of W, and then $(\vec{v}, \vec{v}_2, \ldots, \vec{v}_n)$ is an orthogonal basis of V.

□

Corollary 38.5 *If V is finite-dimensional and* $\mathrm{char}(K) \neq 2$*, then there exists linearly independent linear forms $\varphi_1, \ldots, \varphi_r \in V^*$ and scalars $\lambda_1, \ldots, \lambda_r \in K$ such that, for every $\vec{v} \in V$,*

$$f(\vec{v}, \vec{v}) = \sum_{i=1}^{r} \lambda_i (\varphi_i(\vec{v}))^2.$$

Proof. Let $(\vec{v}_1, \ldots, \vec{v}_n)$ be an orthogonal basis of V, let $(\vec{v}_1^*, \ldots, \vec{v}_n^*)$ be the dual basis, and write $\lambda_i = f(\vec{v}_i, \vec{v}_i)$. By remark 38.2, we have

$$f(\vec{v}, \vec{v}) = \sum_{i=1}^{n} \lambda_i (\vec{v}_i^*(\vec{v}))^2$$

for every $\vec{v} \in V$.

□

39 Real vector spaces

Now assume that $K = \mathbb{R}$ and that V is finite-dimensional.

Theorem 39.1 *Let f be a symmetric bilinear form on V. Then there exists nonnegative integers p and q and linearly independent linear forms $\varphi_1, \ldots, \varphi_{p+q} \in V^*$ such that, for every $\vec{v} \in V$,*

$$f(\vec{v}, \vec{v}) = \sum_{i=1}^{p} (\varphi_i(\vec{v}))^2 - \sum_{j=1}^{q} (\varphi_{p+j}(\vec{v}))^2.$$

Moreover, the integers p and q are uniquely determined by f, we have $\dim(\mathrm{Ker}(f)) = n - (p+q)$ *(so f is non-degenerate if and only if $p + q = \dim(V)$), and f is definite is and only if $p = \dim(V)$ or $q = \dim(V)$.*

Proof. Let's show that existence of p and q and $\varphi_1, \ldots, \varphi_{p+q}$. By corollary 38.5, we have linearly independent linear forms $\chi_1, \ldots, \chi_r \in V^*$ and scalars $\lambda_1, \ldots, \lambda_r \in \mathbb{R}$ such that

$$f(\vec{v}, \vec{v}) = \sum_{i=1}^{r} \lambda_i (\chi_i^*(\vec{v}))^2.$$

Let p be the sunmber of positive λ_i's. After changing the order of the χ_i and deleting the ones for which $\lambda_i = 0$, we may assume that $\lambda_i > 0$ for $1 \leq i \leq p$ and $\lambda_i < 0$ for $p + 1 \leq i \leq r$. For $1 \leq i \leq p$, choose $\mu_i \in \mathbb{R}$ such that $\mu_i^2 = \lambda_i$ and set $\varphi_i = \mu_i \chi_i$. For $r + 1 \leq i \leq r$, choose $\mu_i \in \mathbb{R}$ such that $\mu_i^2 = -\lambda_i$ and set $\varphi = \mu_i \chi_i$. Then these φ_i work.

Now we show the uniqueness of p and q. Suppose that we have two pairs of integers (p, q) and (r, s) and two families $(\varphi_1, \ldots, \varphi_{p+q})$ and $(\psi_1, \ldots, \psi_{r+s})$ satisfying the condition of the theorem, and that $p \neq r$. Without loss of generality, we may assume that $r > p$. Complete $(\psi_1, \ldots, \psi_{r+s})$ to a basis $(\psi_1, \ldots, \psi_n)$ of V^*. Then the family $(\varphi_1, \ldots, \varphi_p, \psi_{r+1}, \ldots, \psi_n)$ has $n - r + p \leq n - 1$ elements, so its span E is not equal to V^*, so there exists $\vec{v} \in V$ nonzero such that

$$\varphi_1(\vec{v}) = \cdots = \varphi_p(v) = \psi_{r+1}(\vec{v}) = \cdots = \psi_n(\vec{v}) = 0.$$

Using the fact that

$$f(\vec{v}, \vec{v}) = \sum_{i=1}^{p} (\varphi_i(\vec{v}))^2 - \sum_{j=1}^{q} (\varphi_{p+j}(\vec{v}))^2,$$

we see that $f(\vec{v}, \vec{v}) \leq 0$. On the other hand, we cannot have $\psi_1(\vec{v}) = \cdots = \psi_r(\vec{v}) = 0$, otherwise $\psi_i(\vec{v})$ would be 0 for every $i \in \{1, \ldots, n\}$, which would force $\vec{v}$ to be 0. So at least one $\psi_i(\vec{v})$ is nonzero for $1 \leq i \leq r$, and using

$$f(\vec{v}, \vec{v}) = \sum_{i=1}^{r} (\psi_i(\vec{v}))^2 - \sum_{j=1}^{s} (\psi_{r+j}(\vec{v}))^2,$$

we see that $f(\vec{v}, \vec{v}) > 0$. This is a contradiction, and so $p = r$. The proof that $q = s$ is similar.

Let's prove the last statement. By proposition 35.4, we have

$$f(\vec{v},\vec{w}) = \frac{1}{2}(f(\vec{v}+\vec{w},\vec{v}+\vec{w}) - f(\vec{v},\vec{v}) - f(\vec{w},\vec{w})).$$

If we write

$$f(\vec{v},\vec{v}) = \sum_{i=1}^{p}(\varphi_i(\vec{v}))^2 - \sum_{j=1}^{q}(\varphi_{p+j}(\vec{v}))^2$$

as above, this gives

$$f(\vec{v},\vec{w}) = \sum_{i=1}^{p}\varphi_i(\vec{v})\varphi_i(\vec{w}) - \sum_{j=1}^{q}\varphi_{p+j}(\vec{v})\varphi_{p+j}(\vec{w}).$$

Complete $(\varphi_1,\dots,\varphi_{p+q})$ to a basis $(\varphi_1,\dots,\varphi_n)$ of V^*, and let $(\vec{v}_1,\dots,\vec{v}_n)$ be the basis of V such that $(\vec{v}_1^*,\dots,\vec{v}_n^*) = (\varphi_1,\dots,\varphi_n)$. I claim that $\mathrm{Ker}(f) = \mathrm{Span}(\vec{v}_{p+q+1},\dots,\vec{v}_n)$ is a basis of $\mathrm{Ker}(f)$, which will prove that $\dim(\mathrm{Ker}(f)) = n-(p+q)$. Indeed, these vectors are clearly in $\mathrm{Ker}(f)$, so we just need to show that they span it. Let $\vec{v}\in\mathrm{Ker}(f)$, and write $\vec{v} = a_1\vec{v}_1+\cdots+a_n\vec{v}_n$. Then, for $1\le i\le p+q$,

$$0 = f(\vec{v},\vec{v}_i) = \pm a_i,$$

hence $a_i = 0$. Now assume that f is definite. It has to be non-degenerate, so $p+q=n$. if $p\ge 1$ and $q\ge 1$, then

$$f(\vec{v}_1+\vec{v}_n,\vec{v}_1+\vec{v}_n) = \varphi_1(\vec{v}_1)^2 - \varphi_n(\vec{v}_n)^2 = 0,$$

contradiction. So $p=0$ or $q=0$. Conversely, if $p=n$, then $f(\vec{v},\vec{v})>0$ for every $\vec{v}\neq 0$, so f is definite. Similarly, if $q=n$, then $f(\vec{v},\vec{v})<0$ for every $\vec{v}\neq 0$, so f is definite.

□

40 Inner products

We still take $K=\mathbb{R}$.

Definition 40.1 We say that a definite symmetric bilinear form on a $\mathbb{R}$-vector space V is *positive* (resp. *negative*) if, for every nonzero $\vec{v}\in V$, $f(\vec{v},\vec{v})>0$ (resp. $f(\vec{v},\vec{v})<0$).

An *inner product* on a $\mathbb{R}$-vector space V is a positive definite symmetric bilinear form on V.

A $\mathbb{R}$-space space together with an inner product on it is called an *inner product space*. We often denote its inner product by $\langle,\rangle$ instead of f, and we write $\|\vec{v}\| = \sqrt{\langle\vec{v},\vec{v}\rangle}$ and call it the *norm* of $\vec{v}$.

A finite-dimensional inner product space is called an *Euclidian space*. An inner product space that is complete (for the distance function given by the inner product) is called a *(real) Hilbert space*.

Example 40.2 $\mathbb{R}^n$ with the usual (standard) inner product :

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle = \sum_{i=1}^{n} x_i y_i$$

(or $\langle \vec{x}, \vec{y} \rangle = \vec{x}^T \vec{y}$ for $\vec{x}, \vec{y} \in M_{n1}(\mathbb{R}) = \mathbb{R}^n$), $M_n(\mathbb{R})$ with $\langle A, B \rangle = \mathrm{Tr}(A^T B)$, $V :=$ {continuous functions $[0,1] \to \mathbb{R}$} with $\langle f, g \rangle \mapsto \int_0^1 f(t)g(t)dt$.

Theorem 40.3 *(Cauchy-Schwarz inequality) Let V be a $\mathbb{R}$-vector space, and let f be a symmetric bilinear form on V such that $f(\vec{v}, \vec{v}) \geq 0$ for every $\vec{v} \in V$.*

Then, for every $\vec{v}, \vec{w} \in V$,

$$f(\vec{v}, \vec{w})^2 \leq f(\vec{v}, \vec{v}) f(\vec{w}, \vec{w}).$$

If moreover f is an inner product (i.e. if f is also definite), then the above inequality is an equality if and only if the family $(\vec{v}, \vec{w})$ is linearly dependent (i.e. the vectors are collinear).

Proof. For every $\lambda \in \mathbb{R}$, we have

$$f(\lambda \vec{v} + \vec{w}, \lambda \vec{v} + \vec{w}) = \lambda^2 f(\vec{v}, \vec{v})^2 + 2\lambda f(\vec{v}, \vec{w}) + f(\vec{w}, \vec{w})^2 \geq 0$$

So the discriminant of this degree 2 polynomial in λ is nonpositive, i.e.

$$f(\vec{v}, \vec{w})^2 - f(\vec{v}, \vec{v}) f(\vec{w}, \vec{w}) < 0.$$

If moreover f is definite and $\vec{v}, \vec{w}$ are linearly independent, then, because $\lambda \vec{v} + \vec{w} \neq 0$ for every λ, the degree 2 polynomial in λ above almost takes positive positive values, so its discriminant is negative, and we get the strict inequality.

□

Let's reformulate this in the inner product case.

Corollary 40.4 *(Cauchy-Schwarz inequality) Let V be an inner product space. Then, for every $\vec{v}, \vec{w} \in V$,*

$$|\langle \vec{v}, \vec{w} \rangle| \leq \|\vec{v}, \vec{v}\| \|\vec{w}, \vec{w}\|,$$

with equality if and only if the family $(\vec{v}, \vec{w})$ is linearly dependent (i.e. the vectors are collinear).

Corollary 40.5 *(Minkowski inequality) Let V be an inner product space. Then, for every $\vec{v}, \vec{w} \in V$,*

$$\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|.$$

Proof. We have

$$\|\vec{v}+\vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2 + 2\langle\vec{v},\vec{w}\rangle$$

by definition, so the Cauchy-Schwarz inequality gives

$$\|\vec{v}+\vec{w}\|^2 \leq \|\vec{v}\|^2 + \|\vec{w}\|^2 + 2\|\vec{v}\|\|\vec{w}\| = (\|\vec{v}\| + \|\vec{w}\|)^2.$$

□

Corollary 40.6 *Let V be an inner product space. Then the function $d : V \times V \to \mathbb{R}_{\geq 0}$ defined by $d(\vec{v},\vec{w}) = \|\vec{v}-\vec{w}\|$ is a distance function on V.*

Proposition 40.7 *Let V be an inner product space. Then, for every $\vec{v},\vec{w} \in V$,*

$$\|\vec{v}+\vec{w}\|^2 + \|\vec{v}-\vec{w}\|^2 = 2(\|\vec{v}\|^2 + \|\vec{w}\|^2).$$

Proof. It's a direct calculation from the definition of $\|.\|$.

□

Proposition 40.8 *(Pythagorean theorem) Let V be an inner product space. Then, if $\vec{v},\vec{w} \in V$ are orthogonal, we have*

$$\|\vec{v}+\vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2.$$

Proof. It's a direct calculation from the definition of $\|.\|$.

□

41 Orthonormal bases

Let V be an inner product space, and assume that it's finite-dimensional.

Remember that, by corollary 37.5, we have $W \oplus W^\perp = V$ and $W = (W^\perp)^\perp$ for every subspace W of V.

Definition 41.1 A basis of V is called *orthonormal* if it is orthogonal and all its elements have norm 1. In other words, if the basis is called $(\vec{v}_1,\ldots,\vec{v}_n)$, then it's orthonormal if

$$\langle\vec{v}_i,\vec{v}_j\rangle = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{otherwise.} \end{cases}$$

Remark 41.2 If $\mathfrak{B}$ is an orthonormal basis of V, then the matrix of the inner product in $\mathfrak{B}$ is I_n. In other words, for every $\vec{v},\vec{w} \in V$,

$$\langle\vec{v},\vec{w}\rangle = [\vec{v}]_{\mathfrak{B}}^T [\vec{w}]_{\mathfrak{B}}.$$

This means that the map $T : \vec{v} \mapsto [\vec{v}]_{\mathfrak{B}}$ is an isomorphism from V to $\mathbb{R}^n$ that sends the inner product of V to the standard inner product of $\mathbb{R}^n$. In particular, T is a homeomorphism of V to $\mathbb{R}^n$.

The coefficients of vectors in an orthonormal basis are particularly easy to calculate :

Proposition 41.3 *Let $\mathfrak{B} = (\vec{v}_1, \dots, \vec{v}_n)$ be an orthonormal basis of V. Then, for every $\vec{v} \in V$, we have*

$$\vec{v} = \sum_{i=1}^{n} \langle \vec{v}, \vec{v}_i \rangle \vec{v}_i.$$

In other words,

$$[\vec{v}]_{\mathfrak{B}} = \begin{pmatrix} \langle \vec{v}, \vec{v}_1 \rangle \\ \vdots \\ \langle \vec{v}, \vec{v}_n \rangle \end{pmatrix}.$$

In particular,

$$\|\vec{v}\|^2 = \sum_{i=1}^{n} \langle \vec{v}, \vec{v}_i \rangle^2.$$

Proof. Let $\vec{v} \in V$, and write $\vec{v} = \lambda_1 \vec{v}_1 + \cdots + \lambda_n \vec{v}_n$. Then, for every $i \in \{1, \dots, n\}$,

$$\langle \vec{v}, \vec{v}_i \rangle = \sum_{j=1}^{n} \lambda_j \langle \vec{v}_j, \vec{v}_i \rangle = \lambda_i.$$

The second formula follows from the Pythagorean theorem (proposition 40.8).

□

By proposition 38.4, we can find an orthogonal basis $(\vec{v}_1, \dots, \vec{v}_n)$ of V, and then $(\|\vec{v}_1\|^{-1}\vec{v}_1, \dots, \|\vec{v}_n\|^{-1}\vec{v}_n)$ is an orthonormal basis.

We will now see an inductive algorithm that takes any basis $(\vec{v}_1, \dots, \vec{v}_n)$ of V and returns an orthogonal basis $(\vec{u}_1, \dots, \vec{u}_n)$ with the property that

$$\mathrm{Span}(\vec{v}_1, \dots, \vec{v}_r) = \mathrm{Span}(\vec{u}_1, \dots, \vec{u}_r)$$

for every $1 \leq r \leq n$. This is called the Gram-Schmidt orthogonalization process.

We can then obtain an orthonormal basis $(\vec{e}_1, \dots, \vec{e}_n)$ by taking $\vec{e}_1 = \frac{1}{\|\vec{u}_i\|}\vec{u}_i$.

Step 1 : Take $\vec{u}_1 = \vec{v}_1$.

Step 2 (nor strictly necessary) : We want to have $\vec{u}_2 = \vec{v}_2 + \lambda_2 \vec{u}_1$ and $\langle \vec{u}_1, \vec{u}_2 \rangle = 0$. This forces us to take

$$\lambda_2 = -\frac{\langle \vec{u}_1, \vec{v}_2 \rangle}{\|\vec{u}_1\|^2}.$$

Induction step : Suppose $\vec{u}_1, \dots, \vec{u}_{r-1}$ constructed, with $r \geq 2$. We want to find $\vec{u}_r$ of the form $\vec{u}_r = \vec{v}_r + \lambda_1 \vec{u}_1 + \cdots + \lambda_{r-1}\vec{u}_{r-1}$, and such that $\langle \vec{u}_r, \vec{u}_i \rangle = 0$ for $1 \leq i \leq r-1$. This forces us to take

$$\lambda_i = -\frac{\langle \vec{u}_i, \vec{v}_r \rangle}{\|\vec{u}_i\|^2}.$$

Here is a consequence of this algorithm :

Proposition 41.4 *Let W be a subspace of V, and let $(\vec{w}_1, \ldots, \vec{w}_r)$ be an orthonormal basis of W. Then we can complete $(\vec{w}_1, \ldots, \vec{w}_r)$ to an orthonormal basis of V.*

Proof. Complete $(\vec{w}_1, \ldots, \vec{w}_r)$ to a basis $(\vec{w}_1, \ldots, \vec{w}_n)$ of V, and apply the Gram-Schmidt algorithm to it. This will not change $\vec{w}_1, \ldots, \vec{w}_r$ (because all the λ_i that appear above will be 0 in the first r steps, as the vectors $\vec{w}_1, \ldots, \vec{w}_r$ are orthogonal), and will output an orthonormal basis of V.

□

Orthogonal projection

Let V be a finite-dimensional inner product space.

Definition 41.5 Let W be a subspace of V. The *orthogonal projection* on W is the endomorphism p of V defined in the following way : For every $\vec{v} \in V$, $p(\vec{v})$ is the unique element of W such that $\vec{v} - p(\vec{v}) \in W^\perp$.

Remark 41.6 We have seen that $V = W \oplus W^\perp$, hence the definition above makes sense and does give a linear transformation. (If $\vec{v} \in V$, then we can write it in a unique way as $\vec{v} = \vec{v}_1 + \vec{v}_2$ with $\vec{v}_1 \in W$ and $\vec{v}_2 \in W^\perp$, and then we take $p(\vec{v}) = \vec{v}_1$. If there were another element $\vec{w}$ of W such that $\vec{v} - \vec{w} \in W^\perp$, then $\vec{v} = \vec{w} + (\vec{v} - \vec{w})$ would be another decomposition of $\vec{v}$ in the direct sum $W \oplus W^\perp$, which is impossible.)

Also, if $\vec{v} \in W$, then $p(\vec{v}) = \vec{v}$. Hence $p(p(\vec{v})) = p(\vec{v})$ for every $\vec{v} \in V$ (because $p(\vec{v}) \in W$) by definition, and p is indeed a projection.

Proposition 41.7 *Let V and W be as above, and let p be the orthogonal projection on W.*

(i) If $(\vec{w}_1, \ldots, \vec{w}_r)$ is an orthonormal basis of W, then, for every $\vec{v} \in V$,

$$p(\vec{v}) = \sum_{i=1}^{r} \langle \vec{v}, \vec{w}_i \rangle \vec{w}_i.$$

(ii) For every $\vec{w} \in W$ such that $\vec{w} \neq p(\vec{v})$, we have

$$\|\vec{v} - \vec{w}\| > \|\vec{v} - p(\vec{v})\|.$$

Part (ii) says that $p(\vec{v})$ is the point of W that is closest to $\vec{v}$. In particular, the distance from $\vec{v}$ to W, defined as

$$d(\vec{v}, W) = \inf_{\vec{w} \in W} d(\vec{v}, \vec{w}),$$

is equal to $\|\vec{v} - p(\vec{v})\|$.

Proof.

(i) By proposition 41.4, we can complete $(\vec{w}_1, \ldots, \vec{w}_r)$ to an orthonormal basis $(\vec{w}_1, \ldots, \vec{w}_n)$ of V. By proposition 41.3, we have $\vec{v} = \sum_{i=1}^{n} \langle \vec{v}, \vec{v}_i \rangle \vec{v}_i$. So, if $\vec{w} = \sum_{i=1}^{r} \langle \vec{v}, \vec{v}_i \rangle \vec{v}_i$, then $\vec{w} \in W$ and $\vec{v} - \vec{w} = \sum_{i=r+1}^{n} \langle \vec{v}, \vec{v}_i \rangle \vec{v}_i \in W^\perp$ (because $\vec{w}_{r+1}, \ldots, \vec{w}_n \in W^\perp$). This implies that $\vec{w} = p(\vec{v})$.

(ii) Let $\vec{v} \in V$ and $\vec{w} \in W$. Then $\vec{w} - p(\vec{v})$ is in $W^\perp$ and $\vec{w} - p(\vec{v})$ is in W, so, by the Pythagorean theorem (proposition 40.8), we have :

$$\|\vec{v} - \vec{w}\|^2 = \|\vec{v} - p(\vec{v})\|^2 + \|p(\vec{v}) - \vec{w}|^2.$$

This shows that $\|\vec{v} - \vec{w}\|$ is always greater than equal to $\|\vec{v} - p(\vec{v})\|$, and is equal to it if and only if $\vec{w} = p(\vec{v})$.

□

42 A little bit of topology

If V is an inner product space, we have seen that we can define a distance function on V by

$$d(\vec{v}, \vec{w}) = \|\vec{v} - \vec{w}\|.$$

We will always use the topology on V given by this distance function.

We will need a few results about this topology.

Proposition 42.1 *Let V be an inner product space. Then the inner product is a continuous map from $V \times V$ to $\mathbb{R}$.*

Proof. Let $\vec{v}_1, \vec{w}_1, \vec{v}_2, \vec{w}_2 \in V$. Then

$$\langle \vec{v}_1, \vec{w}_1 \rangle - \langle \vec{v}_2, \vec{w}_2 \rangle = \langle \vec{v}_1, \vec{w}_1 \rangle - \langle \vec{v}_1, \vec{w}_2 \rangle + \langle \vec{v}_1, \vec{w}_2 \rangle - \langle \vec{v}_2, \vec{w}_2 \rangle = \langle \vec{v}_1, \vec{w}_1 - \vec{w}_2 \rangle + \langle \vec{v}_1 - \vec{v}_2, \vec{w}_2 \rangle.$$

so

$$|\langle \vec{v}_1, \vec{w}_1 \rangle - \langle \vec{v}_2, \vec{w}_2 \rangle| \leq |\langle \vec{v}_1, \vec{w}_1 - \vec{w}_2 \rangle| + |\langle \vec{v}_1 - \vec{v}_2, \vec{w}_2 \rangle| \leq \|\vec{v}_1\| \|\vec{w}_1 - \vec{w}_2\| + \|\vec{v}_1 - \vec{v}_2\| \|\vec{w}_2\|,$$

where the last inequality comes from the Cauchy-Schwarz theorem (theorem 40.3).

Fix $\vec{v}_1$ and $\vec{w}_1$. Let $\delta > 0$ such that $\|\vec{v}_1 - \vec{v}_2\| \leq \delta$ and $\|\vec{w}_1 - \vec{w}_2\| \leq \delta$. Then

$$\|\vec{w}_2\| \leq \|\vec{w}_1\| + \|\vec{w}_2 - \vec{w}_1\| \leq \|\vec{w}_1\| + \delta,$$

so

$$|\langle \vec{v}_1, \vec{w}_1 \rangle - \langle \vec{v}_2, \vec{w}_2 \rangle| \leq \delta \|\vec{v}_1\| + \delta(\|\vec{w}_2\| + \delta).$$

As δ goes to 0, $\delta \|\vec{v}_1\| + \delta(\|\vec{w}_2\| + \delta)$ also goes to 0, and this proves the result.

□

Definition 42.2 Let V be an inner product space. The *unit ball* of V is

$$B = \{\vec{v} \in V | \|\vec{v}\| \leq 1\}.$$

We say that a subset X of V is *bounded* if there exists $\lambda > 0$ such that $X \subset \lambda B$ (i.e. such that $\|\vec{v}\| \leq \lambda$ for every $\vec{v} \in X$).

Theorem 42.3 *Let V and W be* finite-dimensional *inner product spaces. Then any linear transformation $T : V \to W$ is continuous, and moreover if $X \subset V$ is bounded, then so is $T(X) \subset W$.*

Proof. Choose an orthonormal basis $(\vec{e}_1, \ldots, \vec{e}_n)$ of V. (This is not strictly necessary but makes the calculations a bit simpler.) Let $\vec{v}, \vec{w} \in V$, and write $\vec{v} - \vec{w} = \sum_{i=1}^n \lambda_i \vec{e}_i$. Then

$$\|\vec{v} - \vec{w}\|^2 = \sum_{i=1}^n \lambda_i^2.$$

On the other hand, $T(\vec{v} - \vec{w}) = \sum_{i=1}^n \lambda_i T(\vec{e}_i)$, so

$$\|T(\vec{v} - \vec{w})\| \leq \sum_{i=1}^n |\lambda_i| \|T(\vec{e}_i)\| \leq M \sum_{i=1}^n |\lambda_i|,$$

where

$$M = \sup_{1 \leq i \leq n} \|T(\vec{e}_i)\|.$$

If $\|\vec{v} - \vec{w}\| \leq \delta$, then

$$\sum_{i=1}^n \lambda_i^2 \leq \delta^2,$$

so $\|\lambda_i\| \leq \delta$ for every i, and so

$$\|T(\vec{v} - \vec{w})\| \leq nM\delta.$$

As δ goes to 0, $nM\delta$ also goes to 0, so we get the result.

Now let $X \subset V$ be bounded, and suppose that $\|\vec{v}\| \leq \lambda$ for every $\vec{v} \in X$. By the calculation above, if $\|\vec{v}\| \leq \lambda$, then $\|T(\vec{v})\| \leq nM\lambda$. So $T(X)$ is also bounded.

□

Corollary 42.4 *If V is a* finite-dimensional *inner product space, then a subset X of V is compact if and only it is closed and bounded.*

Proof. As V is fnite-dimensional, we have an isomorphism $T : V \to \mathbb{R}^n$, where $n = \dim(V)$. By theorem 42.3, both T and T^{-1} are continuous, so T is a homeomorphism. Moreover, this theorem (applied to T and T^{-1}) also says that $X \subset V$ is bounded if and only if $T(X) \subset \mathbb{R}^n$ is bounded. As we know that a subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded, this gives the result.

□

Remark 42.5 Conversely, if the unit ball of an inner product space V is compact, then V is finite-dimensional. (This is known as Riesz's lemma and is true more generally for any normed vector space, as is theorem 42.3.)

43 Isometries (a.k.a. orthogonal transformations)

Let V be an inner product space.

Definition 43.1 An endomorphism T of V is called an *isometry* (or an *orthogonal transformation*) if, for every $\vec{v} \in V$,

$$\|T(\vec{v})\| = \|\vec{v}\|.$$

Proposition 43.2 *Let $T \in \mathrm{End}(V)$. Then T is an isometry if and only if, for every $\vec{v}, \vec{w} \in V$,*

$$\langle T(\vec{v}), T(\vec{w})\rangle = \langle \vec{v}, \vec{w}\rangle.$$

Proof. If T satisfies the condition of the proposition, then is it an isometry (take $\vec{v} = \vec{w}$). Conversely, suppose that T is an isometry. Let $\vec{v}, \vec{w} \in V$. By proposition 35.4, we have

$$\langle T(\vec{v}), T(\vec{w})\rangle = \frac{1}{2}(\|T(\vec{v}+\vec{w})\|^2 - \|T(\vec{v})\|^2 - \|T(\vec{w})\|^2) = \frac{1}{2}(\|\vec{v}+\vec{w}\|^2 - \|\vec{v}\|^2 - \|\vec{w}\|^2) = \langle \vec{v}, \vec{w}\rangle.$$

□

Definition 43.3 We say that $A \in M_n(\mathbb{R})$ is an *orthogonal matrix* if $AA^T = A^TA = I_n$.

Remark 43.4 (a) Any orthogonal matrix A is invertible, and we have $A^{-1} = A^T$.

(b) By theorem 11.7, $A \in M_n(K)$ is orthogonal if and only if $AA^T = I_n$ if and only if $A^TA = I_n$.

(c) If A is orthogonal, then

$$\det(AA^T) = \det(A)\det(A^T) = \det(A)^2 = \det(I_n) = 1,$$

so $\det(A) = \pm 1$.

(c) I_n is orthogonal.

(d) If $A, B \in M_n(\mathbb{R})$ are orthogonal, then AB and A^{-1} are also orthogonal.

(e) A is orthogonal if and only if A^T is orthogonal.

(f) A is orthogonal if and only if the columns of A form an orthonormal basis of $\mathbb{R}^n$ (for the standard inner product), if and only if the rows of A form an orthonormal basis of $\mathbb{R}^n$.

Theorem 43.5 *Let $T \in \mathrm{End}(V)$, and suppose that V is finite-dimensional. Then the following conditions are equivalent :*

(i) T is an isometry.

(ii) For every orthonormal basis $(\vec{e}_1, \ldots, \vec{e}_n)$ of V, $(T(\vec{e}_1), \ldots, T(\vec{e}_n))$ is an orthonormal basis of V.

(iii) There exists an orthonormal basis $(\vec{e}_1, \ldots, \vec{e}_n)$ of V such that $(T(\vec{e}_1), \ldots, T(\vec{e}_n))$ is an orthonormal basis of V.

(iv) For every orthonormal basis $\mathfrak{B} = (\vec{e}_1, \ldots, \vec{e}_n)$ of V, $[T]_{\mathfrak{B}}$ is an orthogonal matrix.

(v) There exists an orthonormal basis $\mathfrak{B} = (\vec{e}_1, \ldots, \vec{e}_n)$ of V such that $[T]_{\mathfrak{B}}$ is an orthogonal matrix.

In particular, every isometry is an isomorphism and has determinant ± 1.

Proof. It's clear that (ii) implies (iii) and (iv) implies (v).

Let's show that (i) implies (ii). So suppose that T is an isometry, and let $(\vec{e}_1, \ldots, \vec{e}_n)$ be an orthonormal basis of V. By proposition 43.2, we have

$$\langle T(\vec{e}_i), T(\vec{e}_j)\rangle = \langle \vec{e}_i, \vec{e}_j\rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

In particular, the family $(T(\vec{e}_1), \ldots, T(\vec{e}_n))$ is orthogonal, so it's linearly independent by lemma 38.3. As it has n elements, it's a basis of V, and the calculation above says that it's an orthonormal basis.

Let's show that (ii) implies (iv). Let $\mathfrak{B} = (\vec{e}_1, \ldots, \vec{e}_n)$ be an orthonormal basis of V. By (ii), $(T(\vec{e}_1), \ldots, T(\vec{e}_n))$ is also an orthonormal basis of V. By proposition 41.3, we have

$$T(\vec{e}_i) = \sum_{j=1}^{n} \langle T(\vec{e}_i), \vec{e}_j\rangle \vec{e}_j,$$

so the matrix $A = [T]_{\mathfrak{B}}$ is given by

$$A_{ji} = \langle T(\vec{e}_i), \vec{e}_j\rangle.$$

Let's show that $A^T A = I_n$. The (i, j)th entry of $A^T A$ is

$$\sum_{r=1}^{n} (A^T)_{ir} A_{rj} = \sum_{r=1}^{n} A_{ri} A_{rj} = \sum_{r=1}^{n} \langle T(\vec{e}_i), \vec{e}_r\rangle \langle T(\vec{e}_j), \vec{e}_r\rangle.$$

As

$$T(\vec{e}_i) = \sum_{r=1}^{n} \langle T(\vec{e}_i), \vec{e}_r\rangle \vec{e}_r$$

and $\langle .,. \rangle$ is linear in the second variable,

$$(A^T A)_{ij} = \left\langle T(\vec{e}_j), \sum_{r=1}^{n} \langle T(\vec{e}_i), \vec{e}_r \rangle \vec{e}_r \right\rangle = \langle T(\vec{e}_j), T(\vec{e}_i) \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Hence $A^T A = I_n$, which shows that A is orthogonal.

Let's show that (v) implies (iii). If (v) is true, then there exists an orthonormal basis $\mathfrak{B} = (\vec{e}_1, \ldots, \vec{e}_n)$ of V such that the matrix A of T in $\mathfrak{B}$ is orthogonal. Write $A = (a_{ij})$, then we have

$$T(\vec{e}_i) = \sum_{j=1}^{n} a_{ji} \vec{e}_j.$$

So

$$\begin{aligned} \langle T(\vec{e}_i), T(\vec{e}_j) \rangle &= \langle \sum_{r=1}^{n} a_{ri} \vec{e}_r, \sum_{s=1}^{n} a_{sj} \vec{e}_s \rangle \\ &= \sum_{r=1}^{n} \sum_{s=1}^{n} \langle a_{ri} \vec{e}_r, a_{sj} \vec{e}_s \rangle \\ &= \sum_{r=1}^{n} a_{ri} a_{rj} \\ &= (A^T A)_{ij} \\ &= \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

This shows that the family $(T(\vec{e}_1), \ldots, T(\vec{e}_n))$ is orthogonal, so it's linearly independent by lemma 38.3. As it has n elements, it's a basis of V, and the calculation above says that it's an orthonormal basis.

To finish the proof, we just need to show that (iii) implies (i). Let $(\vec{e}_1, \ldots, \vec{e}_n)$ be an orthonormal basis of V such that $(T(\vec{e}_1), \ldots, T(\vec{e}_n))$ is also an orthonormal basis. Let $\vec{v} \in V$. By proposition 41.3,

$$\vec{v} = \sum_{i=1}^{n} \langle \vec{v}, \vec{e}_i \rangle \vec{e}_i,$$

so

$$T(\vec{v}) = \sum_{i=1}^{n} \langle \vec{v}, \vec{e}_i \rangle T(\vec{e}_i).$$

As the $(\vec{e}_1, \ldots, \vec{e}_n)$ and $(T(\vec{e}_1), \ldots, T(\vec{e}_n))$ are both orthonormal, the Pythagorean theorem (proposition 40.8) gives

$$\|\vec{v}\|^2 = \sum_{i=1}^{n} \langle \vec{v}, \vec{e}_i \rangle^2$$

and

$$\|T(\vec{v})\|^2 = \sum_{i=1}^{n} \langle \vec{v}, \vec{e}_i \rangle^2.$$

So $\|\vec{v}\| = \|T(\vec{v})\|$ for every $\vec{v} \in V$, and we have proved that T is an isometry.

□

44 Adjoint of a linear transformation

Let V be an inner product space.

Definition 44.1 Two endomrophisms T_1 and T_2 of V are called *adjoint* if, for every $\vec{v}, \vec{w} \in V$,

$$\langle T_1(\vec{v}), \vec{w} \rangle = \langle \vec{v}, T_2(\vec{w}) \rangle.$$

Note that this definition is symmetric in T_1 and T_2, because the inner product is symmetric.

Proposition 44.2 *Let $T \in \mathrm{End}(V)$. Then there exists at most one $U \in \mathrm{End}(V)$ such that T and U are adjoint.*

If such a U exists, we say that T has an adjoint and we write $U = T^*$ and call it the adjoint of T. Then T^* also has an adjoint, and we have $T = (T^*)^*$.

Proof. Let $U_1, U_2 \in \mathrm{End}(T)$ be such that T and U_1 are adjoint, and that T and U_2 are adjoint. We want to show that $U_1 = U_2$. Let $\vec{v} \in V$. Then, for every $\vec{w} \in V$,

$$\langle \vec{w}, U_1(\vec{v}) - U_2(\vec{v}) \rangle = \langle \vec{w}, U_1(\vec{v}) \rangle - \langle \vec{w}, U_2(\vec{v}) \rangle = \langle T(\vec{w}), \vec{v} \rangle - \langle T(\vec{w}), \vec{v} \rangle = 0.$$

Taking $\vec{w} = U_1(\vec{v}) - U_2(\vec{v})$, we get

$$\|U_1(\vec{v}) - U_2(\vec{v})\|^2 = 0,$$

hence $U_1(\vec{v}) = U_2(\vec{v})$. This shows that $U_1 = U_2$.

If T has an adjoint T^*, then T^* and T are adjoint, so T^* has T as adjoint (i.e. $(T^*)^* = T$).

□

Theorem 44.3 *Suppose that V is finite-dimensional. Then every $T \in \mathrm{End}(V)$ has an adjoint.*

Moreover, for every orthonormal *basis $\mathfrak{B}$ of V,*

$$[T^*]_{\mathfrak{B}} = [T]_{\mathfrak{B}}^T.$$

Proof. Let $T \in \mathrm{End}(V)$. Let $\vec{w} \in V$. The function $\vec{v} \mapsto \langle T(\vec{v}), \vec{w} \rangle$ is a linear transformation on V. As the inner product is definite, it is non-degenerate, and so by (iii)of proposition 36.2 there exists a unique vector $\vec{x} \in V$ such that, for every $\vec{v} \in V$,

$$\langle T(\vec{v}), \vec{w} \rangle = \langle \vec{v}, \vec{x} \rangle.$$

We set $T^*(\vec{w}) = \vec{x}$.

So we have defined a map $T^* : V \to V$, and if we can prove that it is linear, it will be adjoint to T by its definition. Let $\vec{w}_1, \vec{w}_2 \in V$ and $\lambda \in \mathbb{R}$. For every $\vec{v} \in V$, we have

$$\langle \vec{v}, T^*(\lambda\vec{w}_1+\vec{w}_2)\rangle = \langle T(\vec{v}), \lambda\vec{w}_1+\vec{w}_2\rangle = \lambda\langle T(\vec{v}), \vec{w}_1\rangle+\langle T(\vec{v}), \vec{w}_2\rangle = \lambda\langle \vec{v}, T^*(\vec{w}_1)\rangle+\langle \vec{v}, T^*(\vec{w}_2)\rangle$$

$$= \langle \vec{v}, \lambda T^*(\vec{w}_1) + T^*(\vec{w}_2)\rangle.$$

In other words, for every $\vec{v} \in V$,

$$\langle \vec{v}, T^*(\lambda\vec{w}_1 + \vec{w}_2) - \lambda T^*(\vec{w}_1) + T^*(\vec{w}_2)\rangle = 0.$$

Taking $\vec{v} =, T^*(\lambda\vec{w}_1 + \vec{w}_2) - \lambda T^*(\vec{w}_1) + T^*(\vec{w}_2)$, we get that

$$\|T^*(\lambda\vec{w}_1 + \vec{w}_2) - \lambda T^*(\vec{w}_1) + T^*(\vec{w}_2)\|^2 = 0,$$

hence

$$T^*(\lambda\vec{w}_1 + \vec{w}_2) - \lambda T^*(\vec{w}_1) + T^*(\vec{w}_2) = 0.$$

This proves that T^* is linear.

Now we prove the statement about the matrix of T^*. Let $\mathfrak{B}$ be an orthonormal basis of V, and let $A = [T]_{\mathfrak{B}}$. By remark 41.2, we know that, for every $\vec{v}, \vec{w} \in V$,

$$\langle \vec{v}, \vec{w}\rangle = [\vec{v}]_{\mathfrak{B}}^T[\vec{w}]_{\mathfrak{B}}.$$

Now let $\vec{v}, \vec{w} \in V$. We have

$$\langle T(\vec{v}), \vec{w}\rangle = [T(\vec{v})]_{\mathfrak{B}}^T[\vec{w}]_{\mathfrak{B}}(A[\vec{v}]_{\mathfrak{B}})^T[\vec{w}]_{\mathfrak{B}} = [\vec{v}]_{\mathfrak{B}}^T(A^T[\vec{w}]_{\mathfrak{B}}) = \langle \vec{v}, T^*\vec{w}\rangle,$$

so $[T^*(\vec{w})]_{\mathfrak{B}} = A^T[\vec{w}]_{\mathfrak{B}}$. This proves that $[T^*]_{\mathfrak{B}} = A^T$.

□

Remark 44.4 The statement about the matrix of T is totally false if $\mathfrak{B}$ is not an orthonormal basis.

45 The spectral theorem : diagonalization of self-adjoint transformations

Let V be an inner product space.

Definition 45.1 We say that $T \in \mathrm{End}(V)$ is self-adjoint if it is adjoint to itself. In other words, this means that, for every $\vec{v}, \vec{w} \in V$,

$$\langle T(\vec{v}), \vec{w}\rangle = \langle \vec{v}, T(\vec{w})\rangle.$$

Remark 45.2 By theorem 44.3, if V is finite-dimensional and $T \in \mathrm{End}(V)$, then T is self-adjoint if and only if its matrix in some (or any) orthonormal basis is symmetric.

The following lemma is elementary but very useful.

Lemma 45.3 *Let $T \in \mathrm{End}(V)$, and suppose that T is self-adjoint. Let W be a subspace of V such that $T(W) \subset W$.*
Then $T(W^\perp) \subset W^\perp$.

Proof. Let $\vec{v} \in W^\perp$. Then, for every $\vec{w} \in W$,

$$\langle \vec{w}, T(\vec{v}) \rangle = \langle T(\vec{w}), \vec{v} \rangle = 0$$

because $T(\vec{w}) \in W$ and $\vec{v} \in W^\perp$. Hence $T(\vec{v}) \in W^\perp$.

□

Theorem 45.4 *(Spectral theorem.) Suppose that V is finite-dimensional. Let T be a self-adjoint endomorphism of V. Then T is diagonalizable [26] in an orthonormal basis (i.e. there exists an orthonormal basis of eigenvectors of T).*
Here is the matrix version : Let $A \in M_n(\mathbb{R})$ be a symmetric *matrix. Then there exists an orthogonal matrix S such that $SAS^{-1} = SAS^T$ is diagonal.*

The translation betweem the endomorphism and matrix statements uses (f) of remark 43.4.

Proof. By induction on $n := \dim(V)$. The theorem is true for $n = 1$ (because every nonzero vector is an eigenvector of T). Suppose that $n \geq 2$, and that we know the theorem in dimension $n - 1$.

The hardest part is finding the first eigenvectors of T, i.e. showing that T has at least one real eigenvalue. Consider the map $\varphi : V \to \mathbb{R}$ defined by

$$\varphi(\vec{v}) = \langle \vec{v}, T(\vec{v}) \rangle.$$

By proposition 42.1 and theorem 42.3, this is a continuous map. By corollary 42.4, the unit sphere

$$S := \{\vec{v} \in V | \|\vec{v}\| = 1\}$$

of V is compact. Hence there exists $\vec{v}_0 \in S$ such that

$$\lambda := \varphi(\vec{v}_0) = \sup_{\vec{v} \in S} \varphi(\vec{v}).$$

(This λ will be our eigenvalue of T.)

Note also that φ is a quadratic form on V, because T is self-adjoint (so the bilinear form $\langle \vec{v}, T(\vec{w}) \rangle$ is symmetric). Define $\varphi_1 : V \to \mathbb{R}$ by $\varphi_1(\vec{v}) = \lambda \|\vec{v}\|^2 - \varphi(\vec{v})$. This is also a quadratic form on V. For every nonzero $\vec{v} \in V$, we have $\vec{v} = \|\vec{v}\| \vec{x}$ with $\vec{x} = \frac{1}{\|\vec{v}\|} \vec{v} \in S$, so

$$\varphi_1(\vec{v}) = \|\vec{v}\| \varphi_1(\vec{x}) = \|\vec{v}\| (\lambda - \varphi(\vec{x})) \geq 0$$

[26] Over $\mathbb{R}$, so all its eigenvalues are real numbers.

(by definition of λ). Let f_1 be the polar form of φ_1 (proposition 35.4). We have

$$f_1(\vec{v},\vec{w}) = \lambda\langle\vec{v},\vec{w}\rangle - \langle\vec{v},T(\vec{w})\rangle = \langle\vec{v},(\lambda \mathrm{id}_V - T)(\vec{w})\rangle.$$

By the Cauchy-Schwarz inequality (theorem 40.3), we have

$$|f_1(\vec{v},\vec{w})| \leq \varphi_1(\vec{v})\varphi_1(\vec{w})$$

for every $\vec{v},\vec{w} \in V$. As $\varphi_1(\vec{v}_0) = 0$, this implies that $\vec{v}_0 \in \mathrm{Ker}(f_1)$. So $\vec{v}_0$ cannot be in the image of $\lambda\mathrm{id}_V - T$: if $\vec{v}_0 = (\lambda\mathrm{id}_v - T)(\vec{w})$, then

$$0 = f(\vec{v}_0,\vec{w}) = \langle\vec{v}_0,\vec{v}_0\rangle = \|\vec{v}_0\|^2 = 1,$$

contradiction. Hence $\lambda\mathrm{id}_V - T$ is not surjective, so it's not injective, so it has nonzero kernel, so there exists $\vec{e}_1 \in V$ nonzero such that $\lambda\vec{e}_1 = T(\vec{e}_1)$. After dividing $\vec{e}_1$ by $\|\vec{e}_1\|$, we may assume that $\|\vec{e}_1\| = 1$. This is our first eigenvector.

To use the induction hypothesis, need a subspace W of V such that :

- $V = \mathrm{Span}(\vec{e}_1) \oplus W$.
- $\vec{e}_1 \perp W$.
- $T(W) \subset W$.

We can just take $W = \vec{e}_1^{\perp}$, this will satisfy the third condition thanks to lemma 45.3 above.

□

46 The case of $\mathbb{C}$-vector spaces

All the definitions and results of the theory of inner product spaces also work for $\mathbb{C}$-vector spaces (with very similar proofs), but we need to make the followin adaptations (here V is a $\mathbb{C}$-vector space) :

- For every matrix $M \in M_{pq}(\mathbb{C})$,
$$M^* = \overline{M}^T$$
(and $\overline{M}$ is the $p \times q$ matrix with (i,j)th entry $\overline{M}_{ij}$).
We say that the matrix M is *Hermitian* if $M = M^*$. (This is only possible if M is square).

- Instead of bilinear forms, we use *sesquilinear* forms. Those are maps $f : V \times V \to \mathbb{C}$ that are linear in the second variable, compatible with addition in the first variable, and satisfy
$$f(\lambda\vec{v},\vec{w}) = \overline{\lambda} f(\vec{v},\vec{w})$$
for every $\vec{v},\vec{w} \in V$ and $\lambda \in \mathbb{C}$.

- If V is finite-dimensional, the matrix B of a sesquilinear form f in a basis $\mathfrak{B}$ is given by the same formula as in the bilinear case, and we have

$$f(\vec{v}, \vec{w}) = [\vec{v}]_{\mathfrak{B}}^* B [\vec{w}]_{\mathfrak{B}}.$$

- The change of basis formula takes the form $A = P^* B P$ (same notation as in proposition 34.5.

- The condition of being symmetric is replaced by the condition of being *skew-symmetric*, where f skew-symmetric means that

$$f(\vec{w}, \vec{v}) = \overline{f(\vec{v}, \vec{w})}.$$

  In finite dimension, a sesquilinear form is skew-symmetric if and only its matrix in some basis is Hermitian.

- The definition of the kernel and of the isotropic cone don't change.

- If f is a skew-symmetric sesquilinear form on V, then for every $\vec{v} \in V$,

$$f(\vec{v}, \vec{v}) = \overline{f(\vec{v}, \vec{v})},$$

  so $f(\vec{v}, \vec{v}) \in \mathbb{R}$ and it makes sense to say that $f(\vec{v}, \vec{v})$ is positive or negative. We say that f is *definite positive* if $f(\vec{v}, \vec{v}) > 0$ for every nonzero $\vec{v} \in V$.

  Also, we can define Hermitian quadratic forms (analogously to quadratic forms) and relate them to skew-symmetric sesquilinear forms as in the real case. (Note that these forms take values in $\mathbb{R}$, not $\mathbb{C}$.)

- In the analogue proposition 36.2(iii), the map u_f is not linear anymore. It is compatible with addition but we have $u_f(\lambda \vec{v}) = \overline{\lambda} u_f(\vec{v})$. However, u_f is still $\mathbb{R}$-linear, so we can define $\mathrm{Ker}(u_f)$ as in the bilinear symmetric and it is still true that $\mathrm{Ker}(u_f) = \mathrm{Ker}(f)$, and that u_f is injective if and only if f is non-degenerate.

- Everything about orthogonals and orthogonal bases stays true.

- Theorem 39.1 on the signature stays true, but we have to replace the $(\varphi_i(\vec{v}))^2$ by $\|\varphi_i(\vec{v})|^2$.

- A *Hermitian inner product* on V is a positive definite skew-symmetric sesquilinear form on V. We can use this form to define a norm $\|.\|$ and a distance function on V as in the real case. The Cauchy-Schwarz inequality and the various results about the norm are still true.

- A finite-dimensional $\mathbb{C}$-vector space with a Hermitian inner product is called a *Hermitian space*. A $\mathbb{C}$-vector with a Hermitian inner product that makes it complete (as a metric space) is called a *(complex) Hilbert space*.

- The definitions and results about orthonormal bases and the Gram-Schmidt algorithm still work. No modification in the Gram-Schmidt algorithm, but in remark 41.2 we must use $[\vec{v}]^*_{\mathfrak{B}}[\vec{w}]_{\mathfrak{B}}$ (instead of $[\vec{v}]^T_{\mathfrak{B}}[\vec{w}]_{\mathfrak{B}}$), and the coefficients of $\vec{v}$ in an orthonormal basis $(\vec{v}_1, \dots, \vec{v}_n)$ are given by $\langle \vec{v}_i, \vec{v} \rangle$ (instead of $\langle \vec{v}, \vec{v}_i \rangle$).

- Also, the standard Hermitian inner product on $\mathbb{C}^n$ is

$$\langle (x_1, \dots, x_n), (y_1, \dots, y_n) \rangle = \sum_{i=1}^{n} \overline{x}_i y_i.$$

- Orthogonal projections are defined the same way, but teh formula in proposition 41.7 becomes

$$p(\vec{v}) = \sum_{i=1}^{r} \langle \vec{w}_i, \vec{v} \rangle \vec{w}_i.$$

- The topological results are still true.

- The isometry are now called *unitary transformations*, and we say that $A \in M_n(\mathbb{C})$ is a *unitary matrix* if $AA^* = A^*A = I_n$. In finite dimension, T is an isometry if and only if its matrix in an orthonormal basis is unitary.

- Unitary matrices are still stable by multiplication and inversion, and if A is unitary, then $|\det(A)| = 1$. (Because $\det(A^*) = \overline{\det(A)}$ for every $A \in M_n(\mathbb{C})$.)

- The adjoint of an endomorphism of a space with a Hermitian inner products has the same definition as before. If the space is finite-dimensional, the adjoint of T always exists, and its matrix in an orthonormal basis $\mathfrak{B}$ is given by $[T]^*_{\mathfrak{B}}$. So self-adjoint endomorphisms have Hermitian matrices in orthonormal bases.

- The spectral theorem takes the following form : Let V be a finite-dimensional $\mathbb{C}$-vector with a Hermitian inner product. Let T be a self-adjoint endomorphism of V. Then T is diagonalizable in an orthonormal basis, and all the eigenvalues of T are real numbers.

- The matrix version of the spectral theorem becomes : Let $A \in M_n(\mathbb{C})$ be a *Hermitian* matrix. Then there exists a unitary matrix S such that $SAS^{-1} = SAS^*$ is diagonal with real entries.