Linear Algebra

Stéphane Dupraz

1 Vectors and Vector Spaces

1.1 Vector Spaces

A vector space is a collection of "vectors", an abstract notion to designate pretty much anything: numbers, functions, sequences, etc. This generality is the benefit of the abstractness. If a vector space can be a set of anything, what is required to be a vector space is that some operations on elements be possible: a vector space is defined by the structure between elements rather than what the elements are. You do not have to remember the following definition, but to be specific, here are the 8 axioms—all on the operations between elements—that define a vector space.

Definition 1.1. *A **vector space** (over real numbers $\mathbb{R}$) is:*

- *a set V, whose elements are called **vectors**.*
- *an operation $+ : V^2 \to V$ called **addition** (if u and v are vectors, we note $u + v$).*
- *an operation $. : V \times \mathbb{R} \to V$ called **scalar multiplication** (if v is a vector and λ a real, we note λv).*

that satisfies the following 8 axioms:

1. *Associativity of addition : $(u + v) + w = u + (v + w)$.*
2. *Commutativity of addition : $u + v = v + u$.*
3. *Existence of an identity element of addition: there exists an element noted 0 st. $0 + u = u + 0 = u$.*
4. *Existence of inverse elements of addition: for any $u \in V$, there exists an element noted $-u$ st. $u + (-u) = 0$.*
5. *Distributivity of scalar multiplication with respect to vector addition: $\lambda(u + v) = (\lambda u) + (\lambda v)$.*

6. *Distributivity of scalar multiplication with respect to addition in* $\mathbb{R}$ *:* $(\lambda + \mu)v = (\lambda u) + (\lambda v)$.

7. $1v = v$ *(where 1 is the real number).*

8. $(\lambda\mu)v = \lambda(\mu v)$.

It is also possible to define a vector space over $\mathbb{C}$ instead of $\mathbb{R}$, in which case the definition is the same replacing $\mathbb{R}$ by $\mathbb{C}$. We will do so in some specific situations; unless otherwise mentioned, a vector space will be understood as a vector space over $\mathbb{R}$.

A major example of a vector space is $\mathbb{R}^n$, the set of n-tuples of $\mathbb{R}$. An element v of $\mathbb{R}^n$ is $v = (v_1, ..., v_n)$, where the v_i are called the **components** or **coordinates** of v. The addition and scalar multiplication are defined as component-by-component addition and multiplication. The zero for the addition is the vector whose coordinates are all zero.

But there are many other examples, such as spaces of functions: for instance, the set of functions from $\mathbb{R}$ to $\mathbb{R}$ is a vector space. The addition and scalar multiplication are defined by $(f + g)(x) = f(x) + g(x)$ for all x and $(\lambda f)(x) = \lambda f(x)$ for all x.

1.2 Vector Subspaces

Once we have a vector space, an easy way to build others is to take subsets of a vector space, since both operations will satisfy the axioms over the subset. There is one requirement for a subset of a vector space to be a vector space however: when summing two elements of the subset, or multiplying one element of the subset with a scalar, the result must belong to the subset, a property called closure. This defines a vector subspace.

Definition 1.2. *A* ***vector subspace*** W *of the vector space* V *is a subset of* V *such that:*

1. W *contains* 0.

2. W *is* ***close*** *under addition:* $\forall (u, v) \in W^2, u + v \in W$.

3. W *is* ***close*** *under scalar multiplication:* $\forall u \in W, \forall \lambda \in \mathbb{R}, \lambda u \in W$.

A vector subspace is a vector space.

(Note that we could replace the first condition by: W is not empty).

Proof. The proof that W is a vector space consists in painfully checking the 8 axioms. We admit the result. ☐

Vector subspaces can be generalized in a useful way: an affine subspace is the translation of a vector subspace.

Definition 1.3. *A subset A of a vector space V is a an* ***affine subspace of*** *V iff it is of the form $A = v^* + W$, ie. $A = \{v^* + v, v \in W\}$, where v^* is any element of V and W is a vector subspace of V.*

A vector space is an affine space with $v^* = 0$. Affine spaces are particularly useful in geometric applications. For instance, in the plane $\mathbb{R}^2$, a straight line that goes through 0, $\{(x,y)/y = ax\}$, is a vector space. And a straight line $\{(x,y)/y = ax + b\}$ is the translation by (0,b) of the line $\{(x,y)/y = ax\}$ that goes through zero: it is an affine space.

We state one property of vector subspaces: intersecting two vector subspaces, we end up with a vector subspace. Actually, it can even be any intersection (possibly infinite) of vector subspaces.

Proposition 1.1. *Any intersection (possible infinite) of vector subspaces is a vector subspace.*

To prove it, simply check the three definition axioms, as you are asked to do in the problem set.

1.3 Spans and Linear combinations

How to construct vector subspaces? Not any subset A of a vector space V is a vector subspace, far from it. But intuitively, by adding a few elements to S, we may end up with a vector subspace. Actually it is obvious that we can: we just need to add all the vectors of V to end up with V itself. What we would like to do is to add as few vectors as necessary to turn S into a vector subspace, so as to end up with the *smallest* (in the sense of the inclusion) vector subspace that contains S. But are we sure such a smallest vector subspace exists? The inclusion does not define a total order, so if we manage to build W and W' two vector subspaces that contain S, with neither $W \subseteq W'$ nor $W' \subset W$, which one should we decide is the smaller? Luckily, we do not need to answer this question: instead, consider the intersection of W and W'. It is a vector subspace by the previous proposition, it contains S, and it is smaller than both W and W', so it is a better candidate for the smallest vector subspace that contains S. Generalizing the argument, the smallest vector space that contains S exists: we can construct it as the intersection of all the vector subspaces that contain S. We call it the vector subspace spanned by S.

Definition 1.4. *Let V be a vector space and S a subset of V.*

The ***vector subspace spanned (or generated) by*** *S, or* ***linear span of*** *S, noted $Span(S)$, is the smallest vector subspace that contains S, that is the intersection of all the subspaces that contain S:*

$$Span(S) = \bigcap \{W, W \text{ a vector subspace, and } S \subseteq W\}$$

*We say that a set S **span** the vector subspace W.*

This definition is a bit abstract as it does not describe what $Span(S)$ looks like. Let us come to a more descriptive characterization of the linear span when the set S is finite—a finite collection of vectors $v_1, ..., v_n$. To do this, we define linear combinations.

Definition 1.5. *Let $v_1, ..., v_n$ be n vectors of V.*

*A **linear combination** of $v_1, ..., v_n$ is a vector $\lambda_1 v_1 + ... + \lambda_n v_n$ for n scalars $\lambda_1, ..., \lambda_n \in \mathbb{R}$.*

It is easy to check that the set W of all linear combinations of $v_1, ..., v_n$ is a vector subspace of V. Besides, any vector subspace of V that contains the vectors $v_1, ..., v_n$ needs also contain all linear combinations of $v_1, ..., v_n$, i.e. W (the proof is straightforward by induction). In others words, W is the smallest vector subspace that contains $v_1, ..., v_n$: it is the vector subspace spanned by $v_1, ..., v_n$.

Proposition 1.2. *Let $v_1, ..., v_n$ be n vectors of V.*

The vector subspace spanned by $v_1, ..., v_n$ is the set of all linear combinations of $v_1, ..., v_n$:

$$Span(v_1, ..., v_n) = \{\sum_{i=1}^{n} \lambda_i v_i, (\lambda_i)_i \in \mathbb{R}^n\}.$$

The vectors $v_1, ..., v_n$ span a vector subspace W if any element of W can be written as a linear combination of $v_1, ..., v_n$.

Just a remark: we restricted to finite sets S, but it is not much harder to extend the characterization of linear spans with linear combinations to any subset S. In this case, we show instead that $Span(S)$ is the set of all finite linear combinations of S:

$$Span(S) = \{\sum_{i=1}^{n} \lambda_i s_i, n \in \mathbb{N}, s_i \in V \text{ for all } i = 1, ..., n, (\lambda_i)_i \in \mathbb{R}^n\}.$$

1.4 Linear Independence

Definition 1.6. *Let $v_1, ..., v_n$ be n vectors of the vector space V.*

*They are **linearly dependent** iff there exists a linear combination of the (v_i) with not all coefficients equal to zero, that is equal to zero, i.e. reals $(\lambda_i)_i$ not all equal to 0 such that:*

$$\lambda_1 v_1 + ... + \lambda_n v_n = 0$$

Otherwise they are linearly independent. Because the linear combination with all λ_i nil is always equal to zero, linear independence can be phrased as:

Definition 1.7. *The vectors $v_1, ..., v_n$ are **linearly independent** iff:*

$$\lambda_1 v_1 + ... + \lambda_n v_n = 0 \Rightarrow \forall i, \lambda_i = 0.$$

Note that if one of the v_i is zero, then the family is linearly dependent. An essential property of a linearly independent family of vectors is that any vector that can be written as a linear combination of the vectors can be written as such in a unique way. Precisely:

Proposition 1.3. *Let $v_1, ..., v_n$ be linearly independent elements of a vector space V. Let $(\lambda_i)_i$ and $(\mu_i)_i$ be reals. If:*

$$\lambda_1 v_1 + ... + \lambda_n v_n = \mu_1 v_1 + ... + \mu_n v_n,$$

then for all i, $\lambda_i = \mu_i$.

To prove it, just put all terms on one side and use the definition of independence.

1.5 Basis

Consider a family $(v_1, ..., v_n)$ of vectors of V.

- If $(v_1, ..., v_n)$ span V, there exists a way to write any $v \in V$ as a linear composition of the v_i.
- If $(v_1, ..., v_n)$ are linearly independent, the linear decomposition is unique.

Put together, linear independence and spanning defines a basis,

Definition 1.8. *The vector $(v_1, ..., v_n)$ form a **basis** of the vector space V iff they are linearly independent and span V.*

and provides existence and uniqueness of a decomposition in a basis:

Proposition 1.4. ***Existence and uniqueness of the coordinates in a basis.*** *Let $v_1, ..., v_n$ be a basis of the vector space V. Then, for all $v \in V$, there exist a unique n-uple $(\lambda_1, ..., \lambda_n)$ such that:*

$$v = \lambda_1 v_1 + ... + \lambda_n v_n$$

*The $(\lambda_i)_i$ are called the **coordinates** of v in the basis $v_1, ..., v_n$.*

A central example of a basis is the **canonical basis of** $\mathbb{R}^n$. It is $(e_1, ..., e_n)$, where $e_1 = (1, 0, 0, ..., 0)$, $e_2 = (0, 1, 0, ..., 0)$, etc., so that $x = (x_1, ..., x_n) \in \mathbb{R}$ can be written $x = \sum_{i=1}^{n} x_i e_i$.

1.6 Dimension, rank

Given a basis $(v_1, ..., v_n)$ of V, we cannot remove or add vectors to the basis without losing the basis property:

- If we add a new vector $w \in V$ we lose the linear independence property.

  Proof: since $(v_1, ..., v_n)$ spans V, there exist $(\lambda_i)_i$ such that $w = \lambda_1 v_1 + ... + \lambda_n v_n$, which is a linear combination of $(v_1, ..., v_n, w)$ with not all λ_i nil that is equal to zero.

- If we remove a vector—say v_n—we lose the spanning property.

  Proof: otherwise $(v_1, ..., v_{n-1})$ would be a basis. But then, from the previous item, $(v_1, ..., v_n)$ would not be a basis, as it is obtained by adding the vector v_n to $(v_1, ..., v_{n-1})$.

This gives the intuition that two bases of a vector space need have the same number of elements. Proving it is more difficult though, as it requires considering any two bases. We will admit that:

Theorem 1.1. *Provided a basis of a vector space V exists, any two bases of V have the same numbers of elements. Their common number of elements is called the **dimension** of the vector space, and noted $dim(V)$.*

A consequence: if we know that the dimension of a vector space V is m, in order to prove that $(w_1, ..., w_m)$ is a basis of V, we only need to show linear independence. Some vocabulary:

- Note that the theorem proves the uniqueness of a dimension but not its existence: a vector space may not have a finite dimension, i.e. there might exist no basis of a finite number of elements. In this case, we say it is **infinite-dimensional**.

- A vector space can be 0 alone. In this case, we say that its dimension is 0.

- A vector space of dimension 1 is called a **(vector) line**; an affine space of dimension 1 is called an **affine line** (in the plane $\mathbb{R}^2$, they correspond to straight lines).

- A vector space of dimension 2 is called a **(vector) plane**; an affine space of dimension 2 is called an **affine plane**.

- If a vector space V has dimension n, a vector subspace of V of dimension $n-1$ is called a **(vector) hyperplane**; an affine subspace of V of dimension $n-1$ is called an **affine hyperplane**.

We saw that one way to create a vector subspace is to consider the space $Span(v_1, ..., v_n)$ spanned by a family $(v_1, ..., v_n)$ of vectors. The dimension of such vector spaces defines the rank of a family of vectors.

Definition 1.9. *The* ***rank*** $rank(v_1, ..., v_n)$ *of a family of vectors* $v_1, ..., v_n$ *is the dimension of the vector subspace it spans* $dim(Span(v_1, ..., v_n))$.

If r is the rank of a family of n vectors, then $r \leq n$, and $r = n$ iff the vectors are linearly independent.

2 Norms and Inner Products

It is possible to put more structure on a vector space by defining a norm or an inner product on it.

2.1 Norms

The notion of norm generalizes the absolute value of $\mathbb{R}$, $|x| = \max(x, -x)$, to vector spaces:

Definition 2.1. *Let V be a vector space. A **norm** on V is a function $||.||$ from V to $\mathbb{R}+$ (positive reals) that satisfies the following 3 properties:*

1. *Positive-definiteness: If $v \in V$, $||v|| = 0$ iff $v = 0$.*
2. *$\forall v \in V$ and $\lambda \in \mathbb{R}$, $||\lambda v|| = |\lambda|||v||$, where $|\lambda|$ is the absolute value of λ.*
3. *Triangle inequality: $\forall u, v \in V$, $||u + v|| \leq ||u|| + ||v||$.*

*A vector space V endowed with a norm $||.||$ is a **normed vector space**. Rigorously, we note it $(V, ||.||)$ to stress that the norm is part of the definition.*

On $\mathbb{R}$, the absolute value is a norm. On $\mathbb{R}^n$, all the following are norms:

- The L_1 norm: $||v||_1 = \sum_{i=1}^{n} |v_i|$.
- The L_2 norm $||v||_2 = (\sum_{i=1}^{n} |v_i|^2)^{1/2}$.
- The L_p norm: $||v||_p = (\sum_{i=1}^{n} |v_i|^p)^{1/p}$ for any $p \in \mathbb{N}$ more generally.
- The sup norm: $||v||_\infty = \max_{1 \leq i \leq n} |x_i|$.

To prove these are norms, just check the definition axioms. A warning: for the L_p-norm, the triangle inequality is not obvious and is called Minkowski's inequality.

2.2 Inner products

Even more structure is added to a vector space if we define an inner product on it.

Definition 2.2. *Let V be a vector space. An **inner product** of V is a mapping from V^2 to $\mathbb{R}$, noted $< ., . >$: $V^2 \rightarrow \mathbb{R}$, that satisfies the following 3 properties:*

1. *Symmetry:* $\forall u, v \in V, < u, v >=< v, u >$.

2. *Linearity in the first argument:* $\forall u, u', v \in V$, $\forall \lambda, \mu \in \mathbb{R}$, $< \lambda u + \mu u', v >= \lambda < u, v > +\mu < u', v >$.

3. *Positive definiteness:* $\forall u, < u, u >\geq 0$, *and* $< u, u >= 0 \Rightarrow u = 0$.

A vector space and its inner product $(V, < ., . >)$ *is called an* ***inner product space****, or* ***pre-Hilbert space****. An inner product space of finite dimension is called an* ***Euclidean space****.*

Note that symmetry and linearity in the first argument implies linearity in the second argument: an inner product is a **bilinear** mapping. The most common example of an inner product is the **dot product** defined on $\mathbb{R}^n$, which turns $\mathbb{R}^n$ into the most useful Euclidean space. The dot product of two vectors of $\mathbb{R}^n$ $x = (x_1, ..., x_n)'$ and $y = (y_1, ..., y_n)' \in \mathbb{R}^n$ is defined as:

$$x.y = x'y = \sum_{i=1}^{n} x_i y_i$$

A inner product space is more structured than a normed vector space because a norm can always be defined from a inner product.

Proposition 2.1. *Let* $<, >$ *be an inner product over the vector space* V*. Then:*

$$\begin{aligned} ||.|| : V &\to \mathbb{R}+ \\ v &\mapsto \sqrt{< v, v >} \end{aligned}$$

defines a norm on V*.*

Proof. Because $< x, x >\geq 0$ for all x, the function $||x|| = \sqrt{< x, x >}$ is well defined and takes positive values. We can easily check the first two definition axioms of a norm.

1. $||x|| = 0$ iff $< x, x >= 0$, iff $x = 0$ by the positive definiteness of the inner product.

2. $||\lambda x|| = \sqrt{< \lambda x, \lambda x >} = \sqrt{\lambda^2 < x, x >} = |\lambda|||x||$ using the bilinearity of the inner product.

To prove the triangle inequality, we will need Schwarz's inequality, which we will prove in a few paragraphs. So we delay the end of the proof until then. (You can check that none of the results we derive until then rely on the triangle inequality for $||.||$). □

As an example, $||.||_2$ on $\mathbb{R}^n$ is the norm generated by the dot product on $\mathbb{R}^n$. It is why it is also called the **Euclidean norm**. Many norms are not generated by an inner product however. From now on, whenever we are in an inner product space, the norm we use is understood to be the one induced by the inner product, except otherwise mentioned.

2.3 Orthogonality

Once we have an inner product, we can talk about orthogonality.

Definition 2.3. *Two vectors u and v of V are **orthogonal** or **perpendicular** iff their inner product is nil $< u, v >= 0$. We write $u \perp v$.*

Proposition 2.2. ***Pythagore theorem***

Let $(V, <,>)$ be an inner product and $||.||$ the norm induced by $<,>$. Then:

$$u \perp v \Leftrightarrow ||u + v||^2 = ||u||^2 + ||v||^2.$$

Proof. Just use bilinearity and symmetry to show that: $||u + v||^2 =< u + v, u + v >=< u, u > + < u, v > + < v, u > + < v, v >= ||u||^2 + 2 < u, v > +||v||^2$. □

Proposition 2.3. *Let $(v_1, ..., v_m)$ be a family of non-zero vectors of an inner product space $(V, <,>)$. If the v_i are pairwise orthogonal $v_i \perp v_j$ for all $i \neq j$, then the v_i are linearly independent.*

Proof. Let $\sum_{i=1}^{m} \lambda_i v_i = 0$ be a linear combination of the v_i equal to zero. Fix a k between 1 and m and take the inner product between v_k and the linear combination. We get $\lambda_k ||v_k||^2 = 0$. Since $v_k \neq 0$, $\lambda_k = 0$. QED. □

It follows that a family of n pairwise orthogonal vectors forms a basis of a space of dimension n. Such a basis is called an **orthogonal basis**. If in addition, the vectors of the basis are **normed**, in the sense that $||v_k|| = 1$, it is called an **orthonormal basis**. For instance, the canonical basis of $\mathbb{R}^n$ is an orthonormal basis.

Given a non-zero vector u, it is possible to decompose any vector v between: a component proportional to u, λu with $\lambda \in \mathbb{R}$, and a component orthogonal to u, $v - \lambda u$ with $< v - \lambda u, u >= 0$: $v = \lambda u + (v - \lambda u)$. Because the orthogonality condition is equivalent to $\lambda = \frac{<u,v>}{<u,u>}$, we can even explicit the decomposition.

Definition 2.4. *Let $u \neq 0$. For any v, there exists a unique number λ such that $v - \lambda u$ is orthogonal to u, given by:*

$$\lambda = \frac{< u, v >}{||u||^2}$$

*We call λ the **component** of v along u and λu the **projection** of v along u.*

A consequence is Schwarz's inequality:

Proposition 2.4. ***Schwarz's inequality**. For any u, v in the inner product space V,*

$$| < u, v > | \leq ||u||||v||$$

with equality iff u and v are linearly dependent.

Proof. If $u = 0$, the inequality is $0 \leq 0$, which is true. If $u \neq 0$, project v on u: $v = \lambda u + (v - \lambda u)$. By Pythagore theorem, $||v||^2 = ||\lambda u||^2 + ||v - \lambda u||^2 \geq ||\lambda u||^2$ so $||v|| \geq |\lambda|||u|| = \frac{|<u,v>|}{||u||}$. This is an equality iff $v - \lambda u = 0$, i.e. iff u and v are linearly dependent. □

We can now turn back to the proof of the triangle inequality in the proof of proposition 2.1.

Proof. (continued). Using the bilinearity of the inner product,

$$||x + y||^2 =< x + y, x + y >=< x, x > +2 < x, y > + < y, y >= ||x||^2 + 2 < x, y > +||y||^2.$$

Now use Schwarz's inequality $< x, y >\leq | < x, y > | \leq ||x||||y||$, so that:

$$||x + y||^2 \leq ||x||^2 + 2||x||||y|| + ||y||^2 = (||x|| + ||y||)^2.$$

Taking the square root ends the proof. □

3 Matrices

3.1 Definition

Put bluntly, a matrix is a table of numbers.

Definition 3.1. *An $m \times n$ **matrix** is an array with $m \geq 1$ **rows** and $n \geq 1$ **columns**:*

$$A = (a_{ij})_{ij} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

*where the number a_{ij} in the i^{th} row and j^{th} column is called the ij-**entry** or ij-**component**.*

The set of matrices of size $m \times n$ is noted $\mathcal{M}_{mn}$.

We note A_i the i^{th} row of A and A^j the j^{th} column of A, so that:

$$A = (A^1 ... A^n) = \begin{pmatrix} A_1 \\ ... \\ A_m \end{pmatrix}$$

Here are some particular matrices:

- Real numbers can be seen as a 1×1 matrix.
- A vector of $\mathbb{R}^k$ can be seen as a $k \times 1$ matrix (a column vector) or a matrix of size $1 \times k$ (a row vector). By default a vector is seen as a column vector.
- Matrices with as many rows as columns $m = n$ are called **square matrices**.
- The **zero matrix** of $\mathcal{M}_{mn}$ is the matrix with all entries equal to zero.
- A square matrix A is **diagonal** if all its non-diagonal elements are zero: $a_{ij} = 0$ for all i, j such that $i \neq j$. We note $A = diag(a_{11}, ...a_{nn})$.
- The **unit matrix** of size n is the square matrix of size n having all its components equal to zero except the diagonal components, equal to 1. It is noted I_n.
- A square matrix A is **upper-triangular** if all its elements below its diagonal are nil: $a_{ij} = 0$ for all $i > j$.

- A square matrix A is **lower-triangular** if all its elements above its diagonal are nil: $a_{ij} = 0$ for all $i < j$.

3.2 Operations on matrices

A matrix is just a table with numbers. Just as with vector spaces, the set of matrices becomes interesting once we define operations on it.

3.2.1 Addition and scalar multiplication

First, addition and scalar multiplications—the set of matrices $\mathcal{M}_{mn}$ is going to be a vector space. These two are defined on the set $\mathcal{M}_{mn}$ of matrices of the same size $m \times n$.

Definition 3.2. *Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $m \times n$ matrices, and $\lambda \in \mathbb{R}$.*

- *The sum $A + B$ is the matrix whose ij-entry is $a_{ij} + b_{ij}$.*
- *The scalar multiplication of A by λ, λA is the matrix whose ij-entry is λa_{ij}.*

Simply put, we add matrices componentwise and multiply them by scalars componentwise. Once this structure is defined:

Proposition 3.1. *The space $\mathcal{M}_{mn}$ is a vector space of dimension $m \times n$. Its zero is the zero matrix.*

To show it is a vector space, just check the 8 axioms of definition. To get the dimension, notice that if E_{ij} is the matrix whose entries are all zero except the ij entry which is equal to 1, then $(E_{ij})_{i=1\dots m}^{j=1\dots n}$ is a basis of $\mathcal{M}_{mn}$. It is called the canonical basis of $\mathcal{M}_{mn}$.

3.2.2 Multiplication

We define a third operation: a multiplication, although one with different properties than the one of reals. The multiplication is defined over matrices of different sizes, although sizes need to be **conformable**: the product AB is only defined for matrices such that the number of columns of A is equal to the number of rows of B.

Definition 3.3. *Let $A = (a_{ij})$ be an $m \times n$ matrix and $B = (b_{ij})$ be an $n \times s$ matrix.*

Their product AB is the $n \times s$ matrix whose ij-entry is:

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

The following properties of matrix multiplication are easy to verify from the definition.

Proposition 3.2. *Provided conformable matrices:*

- *The unit matrix is the neutral element of matrix multiplication: if A is $m \times n$, then $I_m A = A I_n = A$.*
- *The zero matrix is* ***absorbant****: $A0 = 0A = 0$.*
- *The multiplications is distributive wrt. the addition: $A(B+C) = AB+AC$ and $(B+C)A = BA+CA$.*
- *The multiplication is associative: $A(BC) = (AB)C$.*
- *$A(\lambda B) = \lambda(AB)$.*

But be careful that, contrary to the multiplication on real numbers:

- The matrix multiplication is in general NOT commutative: in general $AB \neq BA$. Here is a counter-example:

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

- $AB = 0$ does NOT imply that either A or B is zero, as it does for the multiplication of reals.

For a square matrix $n \times n$, we wish to talk about its inverse B, a matrix such that $AB = BA = I_n$. With the multiplication on reals, any non-zero $x \in \mathbb{R}$ has an inverse, i.e. another real y such that $xy = 1$ (you may know it as $1/x$). This is not true for matrices: only some matrices have an inverse.

Definition 3.4. *A square matrix of size n is* ***invertible*** *(or* ***non-singular****) iff there exists a matrix B such that $AB = BA = I_n$. Provided existence, the inverse is unique and noted A^{-1}.*

To prove the uniqueness of the inverse, assume that B and C are two inverses of A. Then $B = BI_n = B(AC) = (BA)C = I_nC = C$, which proves that all inverses of A are equal. Obviously, if B is the inverse of A, then A is the inverse of A, so the inverse of the inverse of A is A itself: $(A^{-1})^{-1} = A$. Besides:

Proposition 3.3. *If $A, B \in \mathcal{M}_{nn}$ are invertible, then so is their product AB and:*

$$(AB)^{-1} = B^{-1}A^{-1}$$

To prove it, just check that the suggested inverse works. Be careful with the inverse of the product: we need to "permute A and B".

For square matrices, it is also possible to define the **repeated products**, or **powers** of a square matrix A.

Definition 3.5. *$A^k = A...A$ taken k times. By definition, $A^0 = I_n$.*

We say that a matrix A is **idempotent** if $A^2 = A$. We say that a matrix A is **nilpotent** if $A^k = 0$ for some integer k.

3.2.3 Transpose and symmetric matrices

Another operation on matrices, although less essential, is the transpose; it takes simply one argument—a matrix—and returns another matrix.

Definition 3.6. *Let $A = (a_{ij})$ be a matrix. The* ***transpose*** *A' (or A^t) of A is the matrix obtained by changing its rows into its columns (and vice versa): $A' = (a_{ji})$.*

Obviously, if we apply the transpose operator twice, we end up back on A: $A'' = A$. Note that a row vector is the transpose of a column vector. The following properties of the transpose are easy to verify from the definition.

Proposition 3.4.

- $(\lambda A)' = \lambda A'$.
- *Transpose of the sum:* $(A + B)' = A' + B'$.
- *Transpose of the product:* $(AB)' = B'A'$.
- $(A^{-1})' = (A')^{-1}$ *(provided the inverse exists).*

Be careful with the transpose of the product: we need to "permute A and B".

The transpose of a square matrix is its symmetric wrt. its diagonal. We call a square matrix symmetric when it is symmetric wrt. its diagonal:

Definition 3.7. *A square matrix A is* ***symmetric*** *iff it is equal to its transpose $A' = A$.*

3.3 Matrix as family of vectors, rank of a matrix

If A is an $m \times n$ matrix, we can see its n columns $A^1, ...A^n$ as n vectors of $\mathbb{R}^m$. Conversely, if A^1,..., A^n are n vectors of $\mathbb{R}^m$, we can see them as the $m \times n$ matrix whose columns are the A^j. For instance, note this very useful way to write a linear combination of the vectors A^j using matrix multiplication (just check the equality entry by entry):

$$\lambda_1 A^1 + ... + \lambda_n A^n = A\lambda \text{ , where } \lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

Since we can look at matrices as a family of vectors, we can also consider the vector space that these vectors span:

Definition 3.8. *If $A = (A^1, ..., A^n)$ is an $m \times n$ matrix, we call the space $Span(A^1, ..., A^n)$ spanned by the columns of A the **column space** (or **image**, noted $Im(A)$) of the matrix A.*

We define the rank of a matrix as a natural extension of the rank of a family of vectors.

Definition 3.9. *The **column rank of a matrix** A is the rank of its column space $rank(A^1, ..., A^n)$.*

We can do with rows what we have done with columns. We can see the n rows of an $m \times n$ matrix as m vectors of $\mathbb{R}^n$. We call the space $Span(A_1, ..., A_m)$ spanned by the row vectors of A the **row space** of matrix A and its rank the **row rank** of the matrix. However, the row space is not as much useful as the column space, and:

Proposition 3.5. *The row rank of a matrix is equal to its column rank.*

Proof. The proof is not difficult, but not very interesting. We admit the result. □

A consequence is that the rank of an $m \times n$ matrix is always smaller than both n and m.

Proposition 3.6. *A square matrix A of size n is invertible iff rank(A)=n.*

Proof. Assume that rank(A)=n, i.e. that the columns of A form a basis of $\mathbb{R}^n$. All vectors of $\mathbb{R}^n$ can be expressed as a linear combination of the columns of A. In particular, the vectors $e_j, j = 1, ..., n$ of the canonical basis of $\mathbb{R}^n$. So for all j, there exist a column vector $B^j \in \mathbb{R}^n$ such that $e_j = AB^j$. Noting $B = (B^1, ..., B^n)$, $I_n = AB$ (just pool the vectors as columns of matrices).

To prove that B is the inverse of A, we also need to show that $BA = I_n$. To do so, note that A' also has rank n, so that by the same reasoning there exists C such that $A'C = I_n$. Taking transpose, $C'A = I_n$. But then $BA = (C'A)(BA) = C'(AB)A = C'A = I_n$.

Conversely, assume that A is invertible. We want to show that the columns of A are linearly independent. Consider a linear combination of the columns of A, $A\lambda$ for some vector $\lambda \in \mathbb{R}^n$, that is equal to zero: $A\lambda = 0$. Premultiplying by A^{-1}, $\lambda = A^{-1}0 = 0$. $\square$

4 Systems of Linear Equations

One major application of vector spaces and matrices is to solve systems of linear equations.

4.1 Definition

Definition 4.1. *A* ***system of*** m ***linear equations in*** n ***unknowns*** $x_1, ..., x_n$ *with coefficients* (a_{ij}) *and* $(b_i)_i$ *is:*

$$\begin{cases} a_{11}x_1 + ... + a_{1n}x_n = b_1 \\ \quad ... \\ a_{m1}x_1 + ... + a_{mn}x_n = b_m \end{cases} \qquad (*)$$

Such a system is said to be ***homogenous*** *if* $b = 0$ *(if all the* b_i *are zero).*

$$\begin{cases} a_{11}x_1 + ... + a_{1n}x_n = 0 \\ \quad ... \\ a_{m1}x_1 + ... + a_{mn}x_n = 0 \end{cases} \qquad (**)$$

Given a non-homogenous system $(*)$*, the homogenous system* $(**)$ *with the same* (a_{ij}) *coefficients is called the* ***homogenous system associated to the system*** $(*)$*.*

We can look at the system of linear equations (*) in alternative ways:

- As the matrix equation $Ax = b$ where $A = (a_{ij}), b = (b_i), x = (x_i)$.
- As the linear combination equation $x_1A^1 + ... + x_nA^n = b$ where $A^1, ..., A^n$ are the columns of A: we look for a linear combination of the columns of A equal to b. The question *"Is there a solution x to the system?"* can equivalently be phrased *"Does b belong to the column space of A?"*

4.2 Existence and uniqueness of solutions

We are interested in the existence and number of solutions to such systems of equations. We first consider the case of a homogenous system, then the general case by reducing it to the homogenous case.

4.2.1 Homogeneous systems of equations

The set of solutions to $Ax = 0$ is called the **kernel** of the matrix A and noted Ker(A). It is easy to check that Ker(A) is a vector subspace of $\mathbb{R}^n$. In particular, a homogeneous system has always zero as a solution, called the **trivial solution**. Hence we want to know whether it has non-trivial solutions—whether the the dimension of the kernel is greater than zero.

Theorem 4.1. *Let $Ax = 0$ be a homogeneous system of m linear equations in n unknowns.*

The set of solutions $Ker(A)$ is a vector subspace of $\mathbb{R}^n$ of dimension n-rank(A).

Proof. You are asked to prove the theorem in the problem-set. ☐

This can be understood quite intuitively. Starting from $\mathbb{R}^n$, each *"new independent equation"* imposes a restriction on the possible x that reduces the dimension of the space of solutions by one dimension. *"new independent equation"* here means an equation that does not simply repeat a restriction already contained in the previous equations. This is why we need to subtract the rank, and not the number of equations, to get the dimension of the kernel. Note the following consequences:

- If $n > m$ (more unknowns than equations), then rank(A) $\leq m < n$. So there exist non-trivial solutions.
- If rank(A)=n, there exists a unique solution (the trivial solution zero).

4.2.2 General systems of equations

The general case—when the right-hand side b is not necessarily 0—can be reduced to the case of homogeneous systems quite easily.

Theorem 4.2. *Let $Ax = b$ a linear system of m linear equations in n unknowns.*

- *If b does not belong to the column space of A, then there is no solution.*
- *If b belongs to the column space of A, then there exists at least one solution. If x^* is one **particular solution**, the set of solutions is the affine space $x^* + Ker(A)$, i.e. $\{x = x^* + y, y \in Ker(A)\}$.*

Proof. We have already seen that there exists a solution if and only if b belongs to the column space of A. Now, if there exists one solution x^*, then x is solution to $Ax = b$ iff $y = x - x^*$ is solution to the corresponding homogeneous system $Ay = 0$; iff $x = x^* + y$, $y \in$ Ker(A). ☐

Do note that:

- In contrast to a homogenous system where 0 is always a solution, there might not exist any solution.
- If A is an invertible square matrix m=n=rank(A), there exists a unique solution for any b: since m=rank(A), Im(A) is the whole space $\mathbb{R}^m$ so $b \in \mathbb{R}^m$; since n=rank(A), Ker(A)={0}. This is no surprise: the solution is $x = A^{-1}b$.

5 Elementary operations and Gauss-Jordan elimination

In this section, we consider a set of useful tricks—elementary operations—and a more organized way of using these tricks—Gauss-Jordan elimination—to determine the rank of a matrix or solve a system of linear equations. (In the next section, we will see that these can also be used to calculate the determinant of a matrix).

5.1 Elementary row operations

Here are three operations that we can apply to matrices:

Definition 5.1. *The 3 elementary row (column) operations are:*

1. *Swap the position of two rows (columns).*
2. *Multiply a row (column) by a nonzero scalar.*
3. *Add a scalar multiple of a row (column) to another.*

Why are these operations of any use? If we apply any of these operation to a matrix A, for sure we do change the matrix A. However, the transformed matrix has some properties in common with the matrix A.

5.2 To find the rank of a matrix

First, consider the rank.

Proposition 5.1. *The rank of a matrix does not change when we apply any of the elementary row or column operations.*

To see it, for instance for column operations, simply note that none of the operations change the vector subspace $Span(A_1, ..., A_n)$, so none changes its dimension. So we can use any elementary operations on A to transform it into a matrix whose rank is easier to determine.

5.3 Gauss-Jordan elimination

Gaussian elimination or **reduction to row echelon form** is a systematic way of applying the elementary row operations to turn a matrix into one whose rank is transparent: a matrix in row echelon form. To define a row echelon form matrix, we first need to define a pivot.

Definition 5.2. *Let A be a matrix. For any row A_i of A that does not consist only of zeros, the left-most non-zero entry is called the* ***pivot*** *(or* ***leading coefficient****) of row A_i.*

Definition 5.3. *A matrix B is in* ***row echelon form*** *iff:*

1. *All zero rows are below all non-zero rows.*

2. *For any non-zero row its leading entry is strictly to the right of the leading entry in the previous row.*

The rank of a row echelon form matrix is transparent to read: it is the number of non-zero rows.

Now that we are convinced of the value of row echelon form matrices, let us see how the Gauss-Jordan elimination transforms any matrix into a row echelon form matrix using elementary row operations. Starting from a matrix A, it consists in the following three steps.

1. Find the leftmost non-zero column of the matrix, say column j.

2. If necessary, swap rows (operation 1) so that the entry $1j$ is not zero. Entry $1j$ is now the pivot of the first row.

3. Make all entries below the pivot of the first row $2j$, ..., nj zero by adding an appropriate multiple of the first row to each (operation 3).

Then repeat on the matrix made of the $n-1$ bottom row, and iterate until the last row. The algorithm makes sure that we end up with a matrix in row echelon form. Note however that if we are only interested in the rank of the matrix and this rank becomes obvious at any time during the process, we can stop there!

5.4 To solve a system of linear equations

To solve a system of linear equation, we do need to go up to the row echelon form—and even further. Besides, the equations might have a non-zero right-hand side b which we cannot just forget about—the solutions do depend on b! So first add b as a new column of A.

Definition 5.4. *Let $Ax = b$ be a system of linear equations, with A an $m \times n$ matrix and b a column vector of size n. The* ***augmented matrix*** *of the system is the $m \times (n+1)$ matrix noted $A|b$ obtained by stacking the column vector b as the $(n+1)^{th}$ column of A.*

We will now consider elementary row operations as applied to the augmented matrix $A|b$. That is, when swapping two rows, we swap the elements of b too, etc. (However, when applying the Gauss-Jordan elimination, we do not treat the $(n+1)^{th}$ column as a column).

Proposition 5.2. *Elementary row operations on the augmented matrix $A|b$ do not change the set of solutions to the equation $Ax = b$.*

This is obvious for operations 1 and 2. As for operation 3, just check that a solution to the old system is a solution to the new one, and a solution to the new one a solution to the old one.

The Gauss-Jordan elimination method is useful to solve the equation, because it creates a lot of zeros in the matrix. But not enough zeros: to solve explicitly for the set of solutions, we are going to make appear even more zeros by reducing it to a reduced row echelon form, complementing the Gauss Jordan algorithm.

Definition 5.5. *A matrix A is in* ***reduced row echelon form*** *iff it is in row echelon form and in addition:*

3. All pivots are equal to 1 and are the only nonzero entry in the column.

(In row echelon form, the entries below pivots were zero; now the entries above pivots too).

The solution of a system where the matrix is in reduced row echelon form is straightforward to read. Consider a $m \times n$ matrix A (not the $m \times (n + 1)$ matrix $A|b$) in reduced row echelon form, and let r be the rank (and number of pivots) of A in the reduced row echelon form.

- If there exist zero rows at the bottom of A ($m > r$):
  - If for one zero row i, $b_i \neq 0$, then there is no solution to the system.
  - If for all zero rows i, $b_i = 0$, then these bottom equations are always satisfied: just ignore them. The system can be seen has the $r \times n$ upper-system. Its solutions is an affine space of dimension $n - r$.

  We now assume that the system has solutions and assume $m = r$ without loss of generality.

- If there is one pivot per column ($r = n$), then the system has a unique solution. Since $A = I_n$, the solution is the $(n + 1)^{th}$ column b.
- If there are columns without pivots ($n > r$), then the system has multiple solutions. If j is one of the $n - r$ columns without pivot, treat x_j as a **free variable**, meaning it can take any value. Use the r equations to express the r remaining variables as function of the free variables.

Here's a complement to the Gauss-Jordan algorithm to reduce the matrix to a reduced row echelon form. Starting from the row echelon form.

4. Make all pivots equal to one by dividing each rows by its pivots (operation 2).

5. Make all entries above pivots zero by adding to each an appropriate multiple of the row that contains the pivot (operation 3). (This is just as step 3, but from bottom to top and right to left, instead of top to bottom and left to right).

6 Trace

To each *square* matrix—the trace is only defined for square matrices—we associate a number called its trace, defined as the sum of its diagonal elements.

Definition 6.1. *Let $A = (a_{ij})$ be a square matrix of size n. The* ***trace*** *of A, noted $tr(A)$, is $tr(A) = \sum_{i=1}^{n} a_{ii}$.*

The trace has the following properties:

Proposition 6.1.

- *The trace is linear: $tr(\lambda A) = \lambda tr(A)$ and $tr(A + B) = tr(A) + tr(B)$.*
- *A matrix and its transpose have the same trace: $tr(A') = tr(A)$.*
- *If AB and BA are square (but not necessarily A and B), $tr(AB) = tr(BA)$.*

The first two are straightforward. You can prove the last one by going back to the definition of the product of matrix.

7 Determinants

7.1 Definition

Determinants are a bit abstract but very efficient way to determine whether a family of vectors is linearly dependent or independent (whether matrices are invertible or non invertible). Consider a *square* matrix $A = (a_{ij})$—determinants are only defined for square matrices—or equivalently the family of its n columns $A^1, ..., A^n$. To each matrix A, we are going to attribute a real number called its determinant, noted $\det(A)$, or $|A|$. Thus, we want to define a function from the set of square matrices (of any size) to $\mathbb{R}$ (equivalently from $(\mathbb{R}^n)^n$ to $\mathbb{R}$). We are going to define the determinant by induction, defining the determinant of a square matrix of size n through the determinants of its square submatrices of size $n - 1$.[1]

Definition 7.1. *Let A be a matrix.*

- *A **submatrix** of A is a matrix obtained by deleting any collection of rows and/or columns.*
- *A **minor** of A is the determinant of a* square *submatrix of A.*

To define determinants, we will need to consider only square submatrices (and the associated minors) of size $n - 1$: those obtained by deleting one row and one column. There are n^2 of them. We will note the square submatrix of size $n - 1$ obtained by deleting row i and column j A_{ij}. The corresponding minor $\det(A_{ij})$ is referred to as the ij-minor of A (sometimes called ij**-first minor** too). Now we are ready to define the determinant by induction.

Definition 7.2. *Let A be a square matrix of size n.*

*The **determinant** $\det(A)$ or $|A|$ of A is the real number defined recursively as:*

- *If $n = 1$ (then A is a real a_{11}), $\det(A) = a_{11}$.*
- $\forall n \geq 2$, $\det(A) = (-1)^{1+1} a_{11} \det(A_{11}) + ... + (-1)^{1+n} a_{1n} \det(A_{1n})$.

The recursive step defines $\det(A)$ from n of its minors—all those obtained by removing the first row, and one of the columns; it is called an **expansion along the first row**. (We will soon see that we can also expand the determinant along any other row, or any column, using a similar formula). Note the $(-1)^{i+j}$ factor in front of each term. This makes it convenient to define cofactors—minors adjusted for the sign.

[1] The are several ways to define the determinants; you may have seen another one. All definitions are equivalent of course, and once we picked one as a definition, the others become characterization theorems.

Definition 7.3. *The term* $(-1)^{i+j}\det(A_{ij})$ *is called the* ij ***cofactor*** *of* A*.*

Using the recursive definition, we can derive an explicit expression of the determinant for $n = 2$ and $n = 3$.

Proposition 7.1.

- $n = 2$. *Let* A *be a* 2×2 *matrix:*

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Then $\det(A) = ad - bc$.

- $n = 3$ ***(Rule of Sarrus)****. Let* A *be a* 3×3 *matrix:*

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Then $\det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{21}a_{32}a_{13} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{32}a_{23}$.

To remember the rule of Sarrus: add the product of the elements of the 3 $NO-SE$ diagonals, and subtract the product of the elements of the 3 $NE - SW$ diagonals. After $n = 3$, the general formula gets too complicated for it to be of any practical use in the general case.

7.2 Properties, practical rules

Instead, to calculate determinants in practice we rely on a few properties.

Theorem 7.1. *Let* n *be an integer and consider the determinant function, taking* n *vectors of* $\mathbb{R}^n$ *as argument* $\det : \mathbb{R}^n \times ... \times \mathbb{R}^n \mapsto \mathbb{R}$

1. $\det(I_n) = 1$.
2. ***Multilinearity****:* $\det$ *is linear with respect to each of its argument:*

$$\forall k = 1...n, \det(A^1, ..., \lambda A^k + \mu A^{k'}, ..., A^n) = \lambda \det(A^1, ..., A^k, ..., A^n) + \mu \det(A^1, ..., A^{k'}, ..., A^n)$$

3. If any two columns of A are equal, then $\det(A) = 0$.

4. ***Antisymmetry****: If two columns of A are interchanged, then the determinant changes by a sign.*

5. If one adds a scalar multiple of one column to another then the determinant does not change.

Proof. We admit 1,2,3 (the proofs are essentially by induction on n; remember that we defined the determinant by induction). We prove 4 and 5 from 2 and 3. First, property 4. Just to ease notations, assume the two interchanged columns are the first and the second ones:

$$\begin{aligned}\det(A^1, A^2, ..., A^n) + \det(A^2, A^1, ..., A^n) &= \det(A^1, A^2, ..., A^n) + \det(A^1, A^1, ..., A^n) \\ &\quad + \det(A^2, A^2, ..., A^n) + \det(A^2, A^1, ..., A^n) \\ &= \det(A^1 + A^2, A^1 + A^2, ..., A^n) \\ &= 0\end{aligned}$$

Now property 5: let us add λ times column 1 to column 2:

$$\begin{aligned}\det(A^1, A^2 + \lambda A^1, ..., A^n) &= \det(A^1, A^2, ..., A^n) + \lambda \det(A^1, A^1, ..., A^n) \\ &= \det(A^1, A^2, ..., A^n)\end{aligned}$$

□

Two remarks on multilinearity:

- Note that multilinearity implies that whenever a column is zero, then the determinant is zero.
- Be careful that multilinearity is *not* linearity: a consequence of multilinearity is that $\det(\lambda A) = \lambda^n \det(A)$; resist the urge of dropping the power on λ!

We will admit the following result:

Proposition 7.2. *A matrix and its transpose have the same determinant:* $\det(A') = \det(A)$.

It follows that properties 2 to 5 also apply to columns. Another generalization concerns expanding a determinant.

Proposition 7.3. *We can extend* $\det(A)$ *along any row or any column (not only the first row), according* *to the formulae:*

$$\det(A) = (-1)^{i+1} a_{i1} \det(A_{i1}) + ... + (-1)^{i+n} a_{in} \det(A_{in}) \textit{(expansion along the } i^{th} \textit{ row)}$$

$$\det(A) = (-1)^{1+j} a_{1j} \det(A_{1j}) + ... + (-1)^{n+j} a_{nj} \det(A_{nj}) \textit{(expansion along the } j^{th} \textit{ column)}$$

Proof. That we can expand along any row follows (not completely trivially) from the fact that swapping row i with the first row only changes the sign of the determinant. That we can expand along any column follows from $\det(A') = \det(A)$. □

Proposition 7.4. *The determinant of a triangular (this includes diagonal) matrix* $A = (a_{ij})$ *is the product of its diagonal elements* $\det(A) = \prod_{i=1}^{n} a_{ii}$.

Proof. The proof is by induction. Prove it as an exercise. □

Proposition 7.5. $\det(AB) = \det(A)\det(B)$

Proof. We *admit* the result. □

Note the consequence: if A is invertible, $\det(A^{-1}) = 1/\det(A)$.

7.3 Determinants and rank/linear independence

Note the parallel between properties 2, 4 and 5 and the elementary row operations. The determinant of a matrix:

1. Changes sign when we swap the position of two rows (property 4, antisymmetry).
2. Is multiplied by $\lambda \neq 0$ when we multiply a row by $\lambda \neq 0$ (property 2, multilinearity).
3. Does not change when we add a scalar multiple of a row to another (property 5).

It directly follows that:

Lemma 7.1. *Let* B *be obtained from* A *by elementary row operations.*

Then, $\det(A) = 0$ *iff* $\det(B) = 0$.

Using this result, Gauss-Jordan elimination allows to show the fundamental result:

Theorem 7.2. *n vectors $A^1, ..., A^n$ of $\mathbb{R}^n$ are linearly independent iff* $\det(A^1, ..., A^n) \neq 0$.

(A matrix A is invertible iff $\det(A) \neq 0$*).*

Proof. First prove the result for a matrix B in row echelon form. Because a square row echelon form matrix is upper-triangular and the determinant of a triangular matrix is the product of its diagonal entries, $\det(B)$ is the product of the diagonal entries of B. The rank r of B is the number of its nonzero rows. If $r = n$ (B is invertible), B has no zero rows and hence no zero diagonal entry so $\det(B) \neq 0$. If $r < n$ (B is not invertible), B has at least one zero entry and $\det(B) = 0$. This proves the result for row echelon form matrices. To extend the result to the general case of any square matrix A, just use the previous lemma. □

7.4 Cramer's rule and adjugate matrix

We know that if a system of n equations and n unknown $Ax = b$ has a unique solution iff the rank of the matrix is n, that is iff $\det(A) \neq 0$. The following theorem makes explicit the unique solution using determinants.

Theorem 7.3. ***Cramer's rule*** *Let $(A^1, ..., A^n)$ be n vectors of $\mathbb{R}^n$ that are linearly independent (*$\det(A) \neq 0$*), $b \in \mathbb{R}^n$, and consider the system of linear equations:*

$$x_1 A^1 + ... + x_n A^n = b$$

The unique solution is given explicitly by:

$$\forall i, x_i = \frac{\det(A^1, ..., b, ..., A^n)}{\det(A)}$$

where b occurs in the i^{th} column instead of A^i.

Proof. If x is solution then (b is in i^{th} position):

$$\begin{aligned}\det(A^1, ..., b, ..., A^n) &= \det(A^1, ..., \sum_{k=1}^{n} x_k A^k, ..., A^n) \\ &= \sum_{k=1}^{n} x_k \det(A^1, ..., A^k, ..., A^n) \\ &= x_i \det(A^1, ..., A^i, ..., A^n) \text{ (all the other terms contain two columns that are equal)}\end{aligned}$$

□

Since calculating determinants is not trivial however, this is not very useful in practice for large matrices.

One application of Cramer's rule is to derive a formula for the inverse of a matrix. Let A be an invertible matrix and C the inverse of A, so that $AC = I_n$. This matrix equality is equivalent to the n systems of linear equations $AC^j = e_j$ where e_j is the j^{th} element of the canonical basis of $\mathbb{R}^n$. We can use Cramer's formula to solve each system: $c_{ij} = \det(A^1, ..., e_j, ..., A^n)/\det(A)$, where e_j is in i^{th} position. We can simplify a bit the numerator. Expanding $\det(A^1, ..., e_j, ..., A^n)$ along the i^{th} column, $\det(A^1, ..., e_j, ..., A^n) = (-1)^{i+j}\det(A_{ji})$, so that $c_{ij} = (-1)^{i+j}\det(A_{ji})/\det(A)$. The numerator is the ji-cofactor. Do notice the permutation of indexes: C is the *transpose* of the matrix of cofactors, divided by $\det(A)$.

Definition 7.4. *The* ***adjugate matrix*** *of A, noted $Adj(A)$, is the transpose of the matrix whose entries are the cofactors of A, $(-1)^{i+j}\det(A_{ij})$.*

The result can therefore be stated as:

Proposition 7.6. *If A is invertible, its inverse is its adjugate matrix divided by its determinant:*

$$A^{-1} = \frac{1}{\det(A)} Adj(A)$$

Just as Cramer's rule to solve linear systems, the formula is not a very useful formula to calculate inverses in practice, but it gives the explicit formula for 2×2 matrices that you should know by heart.

Proposition 7.7. *A 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible iff $\det(A) = ad - bc \neq 0$, and then its inverse is given by:*

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

8 Eigenvalues, Eigenvectors and Diagonalization

There are many ways to motivate the study of eigenvalues, eigenvectors and diagonalization, but here is one you will encounter in macroeconomics. Consider a sequence (x_n) of $\mathbb{R}^n$—a function from $\mathbb{N}$ to $\mathbb{R}^n$—defined recursively through a **difference equation**:

$$\forall k \geq 1, x_k = Ax_{k-1}$$

where A is an $n \times n$ matrix. If we have the first term of the sequence x_0, it is easy to express the unique solution to the difference equation using powers of matrices : $x_k = A^k x_0$. But the powers A^k of A are not very easy to handle in general. They are if A is diagonal: if $A = diag(a_1, ..., a_n)$, then $A^k = diag(a_1^k, ..., a_n^k)$. Now imagine that for a non-diagonal matrix A, we can find a diagonal matrix Λ and an invertible matrix P such that $A = P\Lambda P^{-1}$—we will say in this case that A is diagonalizable. Then $A^k = P\Lambda^k P^{-1}$, where Λ^k is easy to calculate. In this section we aim at finding ways to tell whether a matrix A is diagonalizable, and if so, at finding ways to find the matrices Λ and P.

For reasons that will become clear soon, in this section we are going to consider vectors and matrices over $\mathbb{C}$, and not only $\mathbb{R}$: the entries of the matrices can be complex numbers, and we can multiply matrices by complex—not only real— scalars. Do not be afraid: everything we have derived so far applies, generalized to vectors $v \in \mathbb{C}^n$, matrices $A \in \mathbb{C}^{m\times n}$ and scalars $\lambda \in \mathbb{C}$ (you can check that none of the proofs used the fact that the scalars were real and not complex).

8.1 Eigenvectors and Eigenvalues

Definition 8.1. *Let A be a square matrix.*

*A vector $v \neq 0$ is an **eigenvector** of A iff there exists a scalar $\lambda \in \mathbb{C}$ such that $Av = \lambda v$. We call λ the eigenvalue associated to v.*

*A scalar $\lambda \in \mathbb{C}$ is an **eigenvalue** of A iff there exists a non-zero vector v such that $Av = \lambda v$. We call v an eigenvector associated to λ.*

Note that we require $v \neq 0$ (for $v = 0$, $A0 = \lambda 0$ for any λ). Also note that the eigenvalue associated to a eigenvector is unique, whereas the eigenvector associated to an eigenvalue never is: if $v \neq 0$ is an eigenvector associated to λ, then any vector αv, $\alpha \in \mathbb{C}$ and $\alpha \neq 0$ also is. More generally, it is straightforward to check that:

Proposition 8.1. *The set of eigenvectors associated to an eigenvalue λ, completed with 0, is a vector subspace of $\mathbb{C}^n$ of dimension ≥ 1. It is called the **eigenspace** associated to λ.*

An eigenvector can only belong to one eigenspace. Besides, eigenvectors that belong to different eigenspaces are linearly independent:

Proposition 8.2. *Let $\lambda_1, ..., \lambda_m$ be distinct eigenvalues of A and $v_1, ..., v_m$ associated eigenvectors. Then $v_1, ..., v_m$ are linearly independent.*

Proof. The proof is by induction. You are asked to do it in the problem-set. □

A corollary is that there are at most n eigenvalues. We can actually generalize this result a bit.

Proposition 8.3. *Let $\lambda_1, ..., \lambda_m$ be distinct eigenvalues of A.*
For each $k = 1...m$, let d_k be the dimension and $v_k^1, ..., v_k^{d_k}$ a basis of the eigenspace associated to λ_k.
Then the union of the bases is a family of linearly independent vectors.

Proof. Again, you are asked to prove this result in the problem-set. □

A corollary is that $d_1 + ... + d_k \leq n$, and the union of the bases is a basis of $\mathbb{R}^n$ iff $d_1 + ... + d_k = n$.

8.2 Characteristic polynomial

How to find all eigenvalues and associated eigenvectors of a matrix A?

$$\begin{aligned}
\lambda \text{ is an eigenvalue of } A &\Leftrightarrow \exists v \neq 0 / Av = \lambda v \\
&\Leftrightarrow \exists v \neq 0 / (A - \lambda I_n)v = 0 \\
&\Leftrightarrow Ker(A - \lambda I_n) \neq \{0\} \\
&\Leftrightarrow (A - \lambda I_n) \text{ is not invertible} \\
&\Leftrightarrow \det(A - \lambda I_n) = 0
\end{aligned}$$

Proposition 8.4. *λ is an eigenvalue of A iff* $\det(A - \lambda I_n) = 0$.

This motivates the definition of the characteristic polynomial.

Definition 8.2. *The* ***characteristic polynomial*** *of a matrix A is the function $P : \mathbb{C} \to \mathbb{C}$ defined by:*

$$P(\lambda) = \det(\lambda I_n - A) = \begin{vmatrix} (\lambda - a_{11}) & -a_{12} & ... & -a_{1n} \\ -a_{21} & (\lambda - a_{22}) & ... & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & ... & (\lambda - a_{nn}) \end{vmatrix}$$

An eigenvalue λ is therefore a root of the characteristic polynomial: $P(\lambda) = 0$.

The function P is called a polynomial because it is easy to see that P is a polynomial function of degree n, that is a function of the form:

$$P(\lambda) = c_n \lambda^n + ... + c_1 \lambda + c_0, c_n \neq 0$$

Finding the eigenvalues of A is therefore equivalent to finding the roots of its characteristic polynomial. For instance for $n = 2$, this reduces to a quadratic equation.

Proposition 8.5. *let* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *be a 2×2 matrix. Then its characteristic polynomial is:*

$$\begin{aligned} P(\lambda) &= \lambda^2 - (a + d)\lambda + (ad - bc) \\ &= \lambda^2 - tr(A)\lambda + \det(A) \end{aligned}$$

We can then solve for the roots of this quadratic equation. A quadratic equation always has two roots in $\mathbb{C}$, provided we view the case of one root—when $P(\lambda) = (\lambda - \lambda^*)^2$—as having 2 identical roots. But if one root is not in $\mathbb{R}$, then the other is not too, and there is no root in $\mathbb{R}$. We can extend these results on the existence of roots of polynomials—hence on the existence of eigenvalues—to the general case for any n. Indeed, we know that (we will admit that):

Theorem 8.1. ***Fundamental theorem of algebra***

Let P be a polynomial of degree n, $P(\lambda) = c_n \lambda^n + ... + c_1 \lambda + c_0$, $c_n \neq 0$.

Then P has n roots in $\mathbb{C}$, λ_1,...,λ_n, not necessarily distinct*, such that:*

$$P(\lambda) = c_n (\lambda - \lambda_1)...(\lambda - \lambda_n)$$

Because the λ_k's are not necessarily distinct, we say that P has n roots ***counting multiplicity****.*

The number of times a root λ_k is repeated is called its ***multiplicity****.*

In particular, there always exists at least one root in $\mathbb{C}$ to a polynomial, hence one eigenvalue in $\mathbb{C}$ to a matrix. But as for the $n = 2$ case, there may be no root in $\mathbb{R}$. Also, note that we find back that the number of eigenvalues is at most n.

To sum up, we find the eigenvalues by finding the roots of the characteristic polynomial. How to find the associated eigenvectors then? Just remember that for a given λ, an eigenvector v is a non-zero solution to $(A - \lambda I_n)v = 0$, which is a linear system of equation we have learned how to solve.

8.3 Diagonalization

Even in $\mathbb{C}$, we cannot always find a basis of eigenvectors of A: even though the sum of the multiplicity of all eigenvalues sum to n, the multiplicity of an eigenvalue λ_i needs not be equal to the dimension of its eigenspace d_i. We can thus be in the case where $d_1 + ... + d_k < n$. When we can find a basis of eigenvectors of A, we say that A is diagonalizable.

Definition 8.3. *An $n \times n$ matrix A is* ***diagonalizable*** *(in $\mathbb{C}$) iff the sum of the dimensions of its eigenspaces is equal to n, equivalently iff n of its eigenvectors form a basis of $\mathbb{C}^n$.*

We say that a real matrix is *diagonalizable in* $\mathbb{R}$ if it is diagonalization and its eigenvalues and eigenvectors are real. Note a straightforward result:

Proposition 8.6. *If a square matrix has n distinct eigenvalues, then it is diagonalizable.*

But this is only a sufficient, not necessary condition.

Where does the name "diagonalizable" come from? Consider A diagonalizable, and note $(P^j)_{j=1}^n$ the eigenvectors, and λ_j the associated eigenvalues (possibly repeated). For all j, $AP^j = \lambda_j P^j$, which can be written $AP = P\Lambda$ if P is the matrix whose columns are the P^j, and $\Lambda = diag(\lambda_1, ..., \lambda_n)$. Since the eigenvectors form a basis, P is invertible, so $A = P\Lambda P^{-1}$. Conversely, if we know that there exist P invertible and Λ diagonal such that $A = P\Lambda P^{-1}$, then we know that the λ's are the eigenvalues of A, with the columns of P the associated eigenvectors.

Proposition 8.7. *A square matrix A is diagonalizable iff there exist an invertible matrix P and a diagonal* *matrix* Λ *such that:*

$$P^{-1}AP = \Lambda$$

The diagonal elements of Λ *are then the eigenvalues of* A*, and the columns of* P *the associated eigenvectors.*

A diagonal matrix A is diagonalizable: $P = I_n$ and $\Lambda = A$ works. This shows that the eigenvalues of a diagonal matrix are its diagonal elements (with the vectors of the canonical basis its associated eigenvectors). So A and Λ in the decomposition above have the same eigenvalues.

8.4 Diagonalization of real symmetric matrices

There is an important set of matrices that are guaranteed to be diagonalizable: real symmetric matrices. More strongly, real symmetric matrices are **orthogonally diagonalizable**, meaning that they can be diagonalized in a basis of *orthonormal* eigenvectors (for the dot product in $\mathbb{R}^n$).

A matrix P whose columns form an orthonormal basis is called **orthogonal** (and not orthonormal—yes, it is confusing). This can be characterized with transposed matrices. If $P = (P^1, ..., P^n)$ is a matrix, then the ij-entry of the product $P'P$ is the dot product of P^i and P^j: $(P'P)_{ij} = (P^i)'P^j = P^i.P^j$. Therefore, the columns of P form an orthonormal basis iff $P'P = I_n$, i.e. iff $P' = P^{-1}$.

Theorem 8.2. *Let* A *be a real symmetric matrix. Then*

- *All its eigenvalues are real.*
- A *is orthogonally diagonalizable in* $\mathbb{R}$*: there exists a real diagonal matrix* Λ *and a real orthogonal matrix* P *such that:*

$$P'AP = \Lambda$$

Proof. We *admit* the theorem. □

9 Positive definiteness

9.1 Quadratic Forms in $\mathbb{R}^n$

Definition 9.1. *A quadratic form on $\mathbb{R}^n$ is a function Q: $\mathbb{R}^n \to \mathbb{R}$ such that there exists an $n \times n$ real matrix A such that:*

$$Q(x) = x'Ax = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j$$

$$= \sum_{i=1}^{n} a_{ii}x_i^2 + \sum_{i=1}^{n}\sum_{j<i}(a_{ij} + a_{ji})x_i x_j$$

The matrix A corresponding to a quadratic form is not unique: for instance, note that $x'Ax = x'A'x$ for all x, so that both A and A' give the same quadratic form. However, for any quadratic form Q, there exists a unique *symmetric* matrix A that represents Q. Then:

$$Q(x) = x'Ax = \sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{i=1}^{n}\sum_{j<i} a_{ij}x_i x_j$$

9.2 Positive Definite Quadratic Forms and Matrices

Definition 9.2.

- *A quadratic form Q is* ***positive definite*** *iff $Q(x) > 0$ for all $x \neq 0$.*
- *A quadratic form Q is* ***positive semidefinite*** *iff $Q(x) \geq 0$ for all x.*
- *A quadratic form Q is* ***negative definite*** *iff $Q(x) < 0$ for all $x \neq 0$.*
- *A quadratic form Q is* ***negative semidefinite*** *iff $Q(x) \leq 0$ for all x.*
- *A quadratic form Q is* ***indefinite*** *iff it is neither positive semidefinite nor negative semidefinite, that is if it takes both positive and negative values.*

All five definitions are also used for real square matrices: for instance, a real matrix A (not necessarily symmetric) is positive definite iff $x'Ax > 0$ for all $x \neq 0$.

Obviously, A is negative definite (semidefinite) iff $-A$ is positive definite (semidefinite). An essential example of a positive semi-definite symmetric matrix will be the variance-covariance matrix of a vector-valued random variable.

We now see two characterization of positive definiteness *for symmetric matrices.*

9.2.1 Characterization of definiteness with principal minors

One characterization of definiteness of a matrix A relies on minors of A. Only a particular type of minors are used:

Definition 9.3. *Let A be a real square symmetric matrix.*

- *A* ***principal minor of order r*** *of A is a minor of A obtained by keeping r columns of A, as well as the* "same" *rows of A (if the i^{th} column is kept, so is the i^{th} row). We note Δ_r* a *principal minor of A of order r (the notation does not identify which one).*
- *The* ***leading principal minor of order r*** *of A is the principal minor obtained by keeping the r first columns and r first rows of A. In other words, it is the determinant of the upper-left submatrix of size $r \times r$. We note it D_k.*

They are n leading principal minors, but many more principal minors: $\binom{n}{k}$ for each k.

Proposition 9.1. *Let A be a real symmetric square matrix of size n.*

- *A is* ***positive definite*** *iff all its n leading principal minors are strictly positive: $\forall k = 1, ..., n, D_k > 0$.*
- *A is* ***negative definite*** *iff $\forall k = 1, ..., n, (-1)^k D_k > 0$.*
- *A is* ***positive semidefinite*** *iff all its principal minors are weakly positive: $\forall k, \forall \Delta_k, \Delta_k \geq 0$.*
- *A is* ***negative semidefinite*** *iff $\forall k, \forall \Delta_k, (-1)^k \Delta_k \geq 0$.*

Proof. We *admit* the result. □

Be careful that for semidefiniteness, all principal minors (not only the leading ones) need to be checked.

If $A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$ is a 2×2 matrix, the characterization translates into:

- A is **positive definite** iff $a > 0$ and $ad - b^2 > 0$.
- A is **negative definite** iff $a < 0$ and $ad - b^2 > 0$.
- A is **positive semidefinite** iff $a \geq 0$, $d \geq 0$, and $ad - b^2 \geq 0$.
- A is **negative semidefinite** iff $a \leq 0$, $d \leq 0$, and $ad - b^2 \geq 0$.

Sometimes (it will be our case in equality-constrained optimization later on), we want to characterize matrices A that are "positive definite only along some dimensions", meaning $x'Ax > 0$ for all $x \neq 0$ such that $Bx = 0$, where B is a $m \times n$ matrix. To generalize the previous proposition to this case, we rely on bordered matrices:

Definition 9.4. *Let A be a $n \times n$ matrix and B a $m \times n$ matrix.*

The ***matrix A bordered by B*** *is the $(n+m) \times (n+m)$ matrix:*

$$C = \begin{pmatrix} 0_{m \times m} & B \\ B' & A \end{pmatrix}$$

We state a result for positive (negative) definiteness only (a version for semidefiniteness exists):

Proposition 9.2. *Let A be a real symmetric square matrix of size n, and B be a $m \times n$ matrix of rank m. Without loss of generality, assume the first m columns of B are linearly independent.*

Let C be the matrix A bordered by B,

D_k the leading principal minors of order k of C.

- *$x'Ax > 0$ for all $x \neq 0$ such that $Bx = 0$ iff $\forall k = m+1, ..., n, (-1)^m D_{m+k} > 0$.*
- *$x'Ax < 0$ for all $x \neq 0$ such that $Bx = 0$ iff $\forall k = m+1, ..., n, (-1)^k D_{m+k} > 0$.*

Proof. We *admit* the result. □

9.2.2 Characterization of definiteness with the sign of eigenvalues

Another characterization of definiteness relies on eigenvalues.

Proposition 9.3. *Let A be a real symmetric square matrix of size n.*

- *A is* ***positive definite*** *iff all its eigenvalues are strictly positive.*
- *A is* ***negative definite*** *iff all its eigenvalues are strictly negative.*
- *A is* ***positive semidefinite*** *iff all its eigenvalues are weakly positive.*
- *A is* ***negative semidefinite*** *iff all its eigenvalues are weakly negative.*
- *A is* ***indefinite*** *iff A has both positive and negative eigenvalues.*

Proof. Assume A is positive definite. Let λ be an eigenvalue of A; there exists $x \neq 0$ such that $Ax = \lambda x$. But then $x'Ax = x'\lambda x = \lambda x'x$. Because A is a real symmetric matrix, its eigenvector $x \in \mathbb{R}$, so $x'x = ||x||_2^2 \geq 0$. Since $x \neq 0$, $x'Ax > 0$, so $\lambda > 0$.

Conversely, assume all the eigenvalues of a matrix are strictly positive. Let us first show the result for a diagonal matrix $\Lambda = diag(\lambda_1, ..., \lambda_n)$. For a diagonal matrix, $x'\Lambda x = \sum_{i=1}^{n} \lambda_i x_i^2$. So if all $\lambda_i > 0$, then $x'\Lambda x > 0$ for all $x \neq 0$: A is positive definite. If A is not diagonal, we use the orthogonal diagonalizability of real symmetric matrices to reduce to the case of a diagonal matrix. Since A is symmetric, then it is diagonalizable in an orthogonal basis: there exist Λ diagonal and P orthogonal such that $A = P\Lambda P'$. So $x'Ax = x'P\Lambda P'x = (P'x)'\Lambda(P'x)$. Since P is invertible, the function $x \mapsto P'x$ is a bijection. So $x'Ax > 0$ for all $x \neq 0$ iff $y'\Lambda y$ for all $y \neq 0$. We are back to the case of a diagonal matrix. QED.

The proof is identical for positive semidefiniteness. For negativity of A, just use the positivity of $-A$. □

9.3 LDL decomposition and Choleski decomposition

Proposition 9.4. *Let A be a real symmetric matrix. Then:*

A is positive definite iff there exists a diagonal matrix D and a lower triangular matrix L with only 1 on its diagonal such that $A = LDL'$. The decomposition is unique and is called the ***triangular factorization of** A or **LDL decomposition of** A.*

Proof. We *admit* the result. □

There is an alternative way of writing this decomposition. From the LDL decomposition, define $P = L\sqrt{D}$, where $\sqrt{D} = diag(\sqrt{d_{11}}, ..., \sqrt{d_{nn}})$. P is lower triangular and has positive diagonal entries. It follows that A is positive definite iff there exists a lower triangular matrix with positive diagonal entries P such that $A = PP'$, and this decomposition is unique. The decomposition is called the **Choleski decomposition of** A.