Lecture Notes - Linear Algebra

MA Math Camp 2021
Columbia University

César Barilla*

This Version : August 19, 2021

Contents

*The present lecture notes were largely based on math camp materials from Palaash Bhargava, Paul Koh, and Xuan Li. All errors in this document are mine. If you find a typo or an error, please send me an email at cesar.barilla@columbia.edu.

1 Vectors and Vector Spaces

1.1 Vector Spaces

Definition 1.1. *A triple* $(V, +, \cdot)$ *where*

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

is said to be a (real) ***vector space*** *(or a vector space over* $\mathbb{R}$*) iff it satisfies the following 7 axioms:*

(1) Vector addition is commutative and associative: $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in V$*,* $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ *and* $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$*.*

(2) Existence of an identity element for the addition: there exists a ***zero vector*** *denoted* $\mathbf{0}$ *st.* $\mathbf{0} + \mathbf{u} = \mathbf{u} + \mathbf{0} = \mathbf{u}, \forall \mathbf{u} \in V$*.*

(3) Existence of an inverse element for the addition: for any $\mathbf{u} \in V$*, there exists an* ***additive inverse*** *of* $\mathbf{u}$*, noted* $-\mathbf{u}$*, st.* $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$*.*

(4) Existence of an identity element $1 \in \mathbb{R}$ *for the scalar multiplication:* $1\mathbf{u} = \mathbf{u}$*.*

(5) Mixed associativity: for any $\lambda_1, \lambda_2 \in \mathbb{R}$ *and* $\mathbf{v} \in V$*, we have* $(\lambda_1\lambda_2)\,\mathbf{v} = \lambda_1\,(\lambda_2\mathbf{v})$

(6) Scalar multiplication is distributive w.r.t. vector addition: $\lambda\,(\mathbf{v_1} + \mathbf{v_2}) = \lambda\mathbf{v_1} + \lambda\mathbf{v_2}$*.*

(7) Scalar multiplication is distributive w.r.t. addition in $\mathbb{R}$*:* $(\lambda_1 + \lambda_2)\,\mathbf{v} = \lambda_1\mathbf{v} + \lambda_2\mathbf{v}$*.*

Notice that the two operators $+ : V \times V \to V$ and $\cdot : \mathbb{R} \times V \to V$ for the vector space are different from $+$ and $\cdot$ defined on $\mathbb{R}$. Also, $\mathbf{0}$ in (2) is the **zero vector**, not the neutral element 0 in $\mathbb{R}$, although we usually (ab)use the same notation.

We can show the following results using the 7 axioms of a vector space:

(1) the zero vector $\mathbf{0}$ is unique in a vector space;

(2) the additive inverse of vector $\mathbf{v} \in V$ is unique;

(3) $-\mathbf{v} = (-1)\mathbf{v}$. And therefore we can define vector subtraction by $\mathbf{v_1} - \mathbf{v_2} := \mathbf{v_1} + (-\mathbf{v_2})$;

(4) $0\mathbf{v} = \mathbf{0}, \forall \mathbf{v}$; $\lambda\mathbf{0} = \mathbf{0}, \forall \lambda \in \mathbb{R}$, and that $\lambda\mathbf{v} = 0$ implies either $\lambda = 0$ or $\mathbf{v} = \mathbf{0}$.

These are left as exercises.

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 $\mathbb{C}$ and we have a complex vector space. We will do so in some specific situations; unless otherwise mentioned, a vector space will be understood as a real vector space .

A major example of a n-dimensional real vector space is $\langle \mathbb{R}^n, +, \cdot \rangle$, where the vector addition and scalar multiplication are defined in a component-by-component fashion. An element $\mathbf{v}$ in $\mathbb{R}^n$ is $\mathbf{v} = (v_1, v_2, \ldots, v_n)$, where $v_i \in \mathbb{R}$ are called **components** or **coordinates** of $\mathbf{v}$. The zero vector is the vector whose components are all zero.

But the concept of vector spaces can be much more general than that. For example, the set of functions from X to $\mathbb{R}$ is a vector space, where X is some arbitrary nonempty set. The vector addition is defined by $(f+g)(x) := f(x) + g(x)$, and the scalar multiplication is defined by $(\lambda f)(x) := \lambda f(x)$. The zero vector is the function that is constant at 0. We can verify that this satisfies all the requirements for a vector space

For a vector space $(V, +, \cdot)$, a subset W is called a **vector subspace** of $(V, +, \cdot)$ iff $\left(W, +|_W, \cdot|_W\right)$ is a vector space, where $+|_W$ and $\cdot|_W$ are $+$ and $\cdot$ defined for V restricted in W.

Proposition 1.2. $(W, +, \cdot)$[1] *is a vector subspace of* $(V, +, \cdot)$ *iff:*

(1) W contains the zero vector $\mathbf{0}$.

*(2) W is **closed** under vector addition: $\forall \mathbf{u}, \mathbf{v} \in W, \mathbf{u} + \mathbf{v} \in W$.*

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

Conditions (2) and (3) can be replaced by the following compound condition :

$$\lambda \mathbf{u} + \mu \mathbf{v} \in W, \forall \lambda, \mu \in \mathbb{R}, \forall \mathbf{u}, \mathbf{v} \in W$$

Note that the way that operations $+$ and $\cdot$ are defined for a general vector space (Definition 1.1) has already implied that a vector space is closed under vector addition and scalar multiplication. The only thing that could possibly go wrong for a subset W to be a vector space itself, is that the result of some operation does not belong to W, and the definition of a subspace prohibits this.

Exercise 1.3. *(i) Let I an interval in $\mathbb{R}$. Show that the set of continuous functions from I to $\mathbb{R}$, denoted $\mathcal{C}(I, \mathbb{R})$, is a vector subspace of the set of functions from I to $\mathbb{R}$.*

(ii) Show that the set of real sequences $\mathbb{R}^{\mathbb{N}}$ is a vector space.

(iii) Show that the set :

$$S := \{u_n \in \mathbb{R}^{\mathbb{N}} | \forall n \in \mathbb{N}, u_{n+2} = 3u_{n+1} + 4u_n\}$$

is a vector subspace of $\mathbb{R}^{\mathbb{N}}$.

(iv) Show that :

$$E := \{(x, y, z) \in \mathbb{R}^3, x + 2y + z = 0\}$$

is a vector subspace of $\mathbb{R}^3$.

Any vector subspace must contain the zero vector $\mathbf{0}$. It can be shown that intersection of vector subspaces is still a vector subspace, but union may not.

Proposition 1.4. *Let E a vector space. The intersection of a family of vector subspaces of E is a vector space.*

This allows us to define the concept of span.

Definition 1.5. *Let $(V, +, \cdot)$ be a vector space and S a subset of V. The **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:*

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

Because intersection of vector subspaces is still a vector subspace, $Span(S)$ is a vector subspace, which we call the **vector subspace spanned (or generated) by** S.

[1] Strictly speaking, it should be $\left(W, +|_W, \cdot|_W\right)$.

1.2 Sums of Vector Subspaces

Let $(E, +, \cdot)$ a vector space and F, G two vector subspaces of E. We define $F + G$ as the set formed by adding a vector in each of the subspaces :

$$F + G := \{z \in E | \exists (x, y) \in F \times G, z = x + y\}$$

We can show that this set is itself a vector space, and is indeed the smallest one containing F and G.

Proposition 1.6. *$F + G$ is a vector subspace of E. Furthermore, itis smallest (in the sense of inclusion) vector subset of E containing both F and G.*

Remark 1.7. *It is important to notice that the notation $F + G$ is defined by analogy only and is* not *an actual summation. For one thing, we define it only for vector subspaces (and not arbitrary subset). It is also important not to confuse $F + G$ with $F \cup G$. In particular, $F \cup G$ is not a vector subspace in general.*

This proposition extends to a sum of n vector subspaces

Proposition 1.8. *If $F_1, ..., F_n$ are vector subspaces of E, denote :*

$$\sum_{i=1}^{n} F_i = F_1 + ... + F_n = \{z \in E, \exists (x_1, ..., x_n) \in F_1 \times ... \times F_n, z = x_1 + ... + x_n\}$$

we have that $\sum_{i=1}^{n} F_i$ is a vector subspace of E, furthermore it is the smallest such subspace containing all the F_i.

We prove directly this general version, the previous proposition will therefore be a corollary.

Proof. We first prove that $G := F_1 + ... + F_n$ is a vector subspace. For all i, F_i is a vector subspace of E hence $0_E \in F_i$, hence :

$$0_E = \underbrace{0_E}_{\in F_1} + \underbrace{0_E}_{\in F_2} + ... + \underbrace{0_E}_{\in F_n} \in \sum_{i=1}^{n} F_i$$

Let $z, z' \in G$, $\lambda \in \mathbb{R}$. There exists $(x_1, ..., x_n) \in \prod_{i=1}^{n} F_i$ and $(x'_1, ..., x'_n) \in \prod_{i=1}^{n} F_i$ such that :

$$z = \sum_{i=1}^{n} x_i$$
$$z' = \sum_{i=1}^{n} x'_i$$

Hence :

$$\begin{aligned} \lambda z + z' &= \lambda \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x'_i \\ &= \sum_{i=1}^{n} \underbrace{\lambda x_i + x'_i}_{\in F_i} \end{aligned}$$

So $\lambda z + z' \in G$.

Next, we show that G contains all the F_i. Fix i, take any $x_i \in F_i$, then :

$$x_i = 0_E + ... + x_i + ... + 0_E \in G$$

Hence $F_i \subset G$

Now assume that F is a vector subspace of E containing each F_i. We want to show that $G \subset F$. Let $z \in G$, there exists $(x_1, ..., x_n) \in \prod_{i=1}^n F_i$ such that $z = \sum_{i=1}^n x_i$. Since $F_i \subset F$ for all i, $x_i \in F$ for all i. Since F is a vector subspace, the sum of the x_i has to be in F. Therefore, $G \subset F$. $\square$

We also say that the sum $F + G$ is *direct* if the decomposition is unique, i.e if for all $z \in F + G$ there exists a *unique* pair $x, y \in F \times G$ such that $z = x + y$. We sometimes denote $F \oplus G$ to signify that the sum is direct, and similarly for the sum of n subspaces.

A sufficient property for the sum $F_1 + ... + F_n$ to be direct is given by the following theorem

Theorem 1.9. *Let $F_1, ..., F_n$ vector subspaces of E. The sum $F_1 + ... + F_n$ is direct if and only if for all $(x_1, ..., x_n) \in F_1 \times ... \times F_n$*

$$x_1 + ... + x_n = 0_E \ \Rightarrow\ \forall i = 1, ..., n,\ x_i = 0_E$$

Proof. First assume that the sum is direct. Let $(x_1, ..., x_n) \in F_1 \times ... \times F_n$. Assume that $x_1 + ... + x_n = 0_E$; observing that $0_E = 0_E + ... + 0_E$ with $0_E \in F_i$ for all i, and since the decomposition is unique we get $x_i = 0_E$ for all i.

Now assume the property instead, and show that the sum is direct. Let $(x_1, ..., x_n), (x'_1, ..., x'_n) \in F_1 \times ... \times F_n$ such that :

$$x_1 + ... + x_n = x'_1 + ... + x'_n$$

Then using vector space properties (existence of the inverse for addition) :

$$(x_1 - x'_1) + ... + (x_n - x'_n) = 0_E$$

Hence for all i, $x_i - x'_i = 0_E$, so $x_i = x'_i$. $\square$

We can give another characterization for the sum to be direct.

Proposition 1.10. *If F and G are two vector subspaces of E, $F + G$ is direct if and only if $F \cap G = \{0_E\}$.*

The proof is left as an exercise.

Example 1.11. *Some examples :*

(i) Let F a vector subspace of E. $F + F = F$. Furthermore $F \cap F$, if $F \neq \{0_E\}$ the sum is not direct.

(ii) $F + \{0_E\} = F$

(iii) $F + G = G + F$ (the of vector subspaces is commutative)

This gives the intuitive idea of decomposing elements of a vector space into sums of elements of smaller subspaces, but we might wonder when this decomposition is unique. This will be made more precise using the concept of linear (in)dependence.

1.3 Collections of Vectors : Linear (In)Dependence and Span

Define linear combinations for arbitrary collections of vectors.

Definition 1.12. *Let $\mathbf{v}_1, ..., \mathbf{v}_n$ be n vectors of V. A **linear combination** of $\mathbf{v}_1, ..., \mathbf{v}_n$ is a vector $\lambda_1 \mathbf{v}_1 + ... + \lambda_n \mathbf{v}_n$ for n scalars $\lambda_1, ..., \lambda_n \in \mathbb{R}$.*

This allows us to introduce the concept of linear independence.

Definition 1.13. *In vector space $(V, +, \cdot)$, a finite set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are said to be **linearly independent**, iff the linear combination $\sum_{i=1}^{n} \lambda_i \mathbf{v}_i = \mathbf{0}$ implies $\lambda_i = \mathbf{0}$ for any i. Otherwise, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are said to be **linearly dependent**.*

Clearly, if a vector can be represented by a linear combination of a set of linearly independent vector, then the representation is unique.

Proposition 1.14. *Let $\mathbf{v}_1, ..., \mathbf{v}_n$ be linearly independent elements of a vector space $(V, +, \cdot)$. Let $(\lambda_i)_{i \in N}$ and $(\mu_i)_{i \in N}$ be reals. If:*

$$\lambda_1 \mathbf{v}_1 + ... + \lambda_n \mathbf{v}_n = \mu_1 \mathbf{v}_1 + ... + \mu_n \mathbf{v}_n,$$

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

Proposition 1.15. *In a vector space, $\mathbf{v}_1, ..., \mathbf{v}_n$ are linearly dependent iff $\exists \mathbf{v}_i$ ($i \in \{1, 2, ..., n\} := N$) that can be represented by a linear combination of $(\mathbf{v}_j)_{j \neq i}$.*

The next theorem is a fundamental result about linear dependency.

Theorem 1.16. *In vector space $(V, +, \cdot)$, if $\mathbf{u}_1, \ldots, \mathbf{u}_m$ can be represented by linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_n$, and $m > n$, then $\mathbf{u}_1, \ldots, \mathbf{u}_m$ are linearly dependent.*

Shortly put, if "more" can be represented by "less", then "more" must be linearly dependent. Let's admit this result without providing a proof, but you may adapt the proof of Lang's Theorem 3.1 to prove it.

You can also observe that if a collection of vectors is linearly independent, every subcollection (obtained by removing some vectors from the collection) is also linearly independent. A supercollection (obtained by adding some vectors) is still independent if and only if the added vectors do not belong to the span of the initial collection.

It is straightforward that for a finite collection $S = \{v_1, ..., v_n\}$ of vectors in V, the span of S is simply the set of linear collection of those vectors :

$$Span(S) = \left\{ z \in V, \exists \lambda_1, ..., \lambda_n, \, z = \sum_{i=1}^{n} \lambda_i x_i \right\}$$

Definition 1.17. *We say that a finite collection $S = \{v_1, ..., v_n\}$ of vectors in V spans a subset $A \subset V$ if $Span(S) = A$.*

It is straightforward to observe that if a collection of vectors spans V (the whole space), then every supercollection (obtained by adding vectors) also spans V.

1.4 Basis

The idea behind a basis is to give a unique representation for each vector as a linear combination of a collection of vectors, by combining linear independence and spanning properties.

Definition 1.18. *Let $(V, +, \cdot)$ a vector space. The collection of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ is called a **(finite) basis** of V if :*

1. *$\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent, and*
2. *$\mathbf{v}_1, \ldots, \mathbf{v}_n$ span V, i.e $Span(\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} = V$*

Condition (2) states that all vectors in V can be represented by a linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Note that because of condition (1), we know that the representation in condition (2) is unique. Given a basis $v_1, ..., v_n$, the unique real numbers $\lambda_1, ..., \lambda_n$ such that $x = \lambda_1 v_1 + \ldots + \lambda_n v_n$ are called the **coordinates** of $x \in V$.

Note that a basis of V is not unique. However, two bases must have the same number of vectors. To see this, suppose $\mathbf{u}_1, \ldots, \mathbf{u}_m$ and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are both bases of V, and $m < n$. Then $\mathbf{u}_1, \ldots, \mathbf{u}_m$ can represent $\mathbf{v}_1, \ldots, \mathbf{v}_n$, and therefore $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly dependent by Theorem 1.16. This contradicts the assumption that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form a basis of V.

Therefore, it is without ambiguity to define the **dimension** of V as the number of vectors in its basis, denoted as $Dim\,(V)$.

The next two propositions shows two other equivalent definitions of bases.

Proposition 1.19. *In vector space $(V, +, \cdot)$, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are a basis of V iff they have the maximum number of linearly independent vectors in V.*

In words, the dimension of V is the largest possible cardinality of a linearly independent collection of vectors in V.

Proof. $\Leftarrow$:

Take any $\mathbf{y} \in V$. WTS y can be represented by a linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

Because $\mathbf{v}_1, \ldots, \mathbf{v}_n$ has the maximum number of linearly independent vectors in S, the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and y are linearly dependent. Therefore, they have a non-trivial representation of the zero vector:

$$\sum_{i=1}^{n} \lambda_i \mathbf{v}_i + \mu \mathbf{y} = \mathbf{0}$$

Observe that $\mu \neq 0$, otherwise $\lambda_i = 0$ for all i, and the representation becomes trivial. Therefore,

$$y = \sum_{i=1}^{n} \left(-\frac{\lambda_i}{\mu} \right) \mathbf{v}_i$$

which represents $\mathbf{y}$ as a linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

$\Rightarrow$:

Suppose there exists a set of linearly independent vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ in S with $m > n$. Because $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are a basis of V, the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ can represent $\mathbf{u}_1, \ldots, \mathbf{u}_m$, and therefore $\mathbf{u}_1, \ldots, \mathbf{u}_m$ are linearly dependent by Theorem 1.16. Contradiction. $\square$

Proposition 1.20. *In vector space $(V, +, \cdot)$, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are a basis of V iff they have the minimum number of vectors in V that can represent all vectors in V as their linear combinations.*

In other words, the dimension of V is the smallest possible cardinality of a collection of vector that spans V.

Proof. $\Leftarrow$:

WTS $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent. Suppose not, then we can represent one of them as a linear combination of the others. Without loss of generality, suppose $\mathbf{v}_n$ can be represented by $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$. Because all vectors in V can be represented by $\mathbf{v}_1, \ldots, \mathbf{v}_n$, they can also be represented by $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$, which contradicts the assumption that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ have the minimum number of vectors in V that can represent all vectors in V.

$\Rightarrow$:

Suppose there exists a set of vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ in V that can represent all vectors in S, and $m < n$. Then $\mathbf{u}_1, \ldots, \mathbf{u}_m$ can represent $\mathbf{v}_1, \ldots, \mathbf{v}_n$, and therefore $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly dependent by Theorem 1.16. This contradicts the assumption that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are a basis of V. $\square$

If $\dim V = n$, then we say that the vector space is **n-dimensional**. Clearly, any set of more than n vectors in an n-dimensional vector space must be linearly dependent.

More specifically, the two results give that a collection that is linearly independent (resp. spans V) has at most (resp. at least) n elements, and has exactly n if and only if it is a basis. We also have reduction and completion results :

Proposition 1.21. *From every collection that spans V, we can extract a basis. Every linearly independent collection in V can be completed into a basis.*

Since any vector subspace of V is also a vector space, we can talk about bases for vector subspaces as well. Sometimes, the dimension of a vector subspace is called the rank to differentiate it from the dimension of the whole ambient space, but it is not necessary to make this distinction. If V is an n-dimensional vector space, a direct consequence of the definition is that the span of S is the same as the span of its basis for any $S \subset V$

Proposition 1.22. *In vector space $(V, +, \cdot)$, suppose $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are a basis of $S \subset V$. then :*

$$Span(S) = Span(\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\})$$

This gives us immediate inequalities in terms of dimension :

Proposition 1.23. *If S is a vector subspace of V, then $dim(S) \leq dim(V)$ and $dim(S) = dim(V)$ if and only if $S = V$.*

More generally, we can prove that any two vector spaces are isomorphic (there exists a linear bijection between them) if and only if they have the same dimension.

We can also relate dimension to direct sum properties.

Proposition 1.24. *If F and G are two vector subspaces of V such that the sum $F \oplus G$ is direct, then :*

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

The so-called Grassman Formula gives a result for general sums of spaces :

Proposition 1.25. *If F and G are two vector subspaces of V, then :*

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

As a result, we also have the following equivalence relating decomposition of a vector space into subspaces and dimension :

Proposition 1.26. *If F and G are two vector subspaces of V, then the following statements are equivalent :*

(i) $V = F \oplus G$

(ii) $dim(V) = dim(F) + dim(G)$ *and* $F \cap G = \emptyset$

(iii) $dim(V) = dim(F) + dim(G)$ *and* $F + G = V$

For example, the space $\mathbb{P}_n$ of all polynomials of degree at most n, consisting of all polynomials of the form

$$p(t) = a_0 + a_1 t + ... + a_n t^n;$$

is an $n+1$-dimensional vector space, for which one basis is $\{1, t, ..., t^n\}$.

It is also possible that a vector space V does not have a finite basis, in which case we say that V is **infinite-dimensional**.

The **canonical basis** of the vector space $\mathbb{R}^n$ is $\mathbf{e}_1, \ldots, \mathbf{e}_n$, where $\mathbf{e}_1 := (1, 0, 0, \ldots, 0)$, $\mathbf{e}_2 := (0, 1, 0 \ldots, 0)$, and so forth. Therefore $\dim \mathbb{R}^n = n$.

1.5 Inner Products, Norms, and Metrics

Let's define an inner product operator on a real vector space V, to give it more structure.

Definition 1.27. *Let $(V, +, \cdot)$ be a vector space. The 4-tuple $\big(V, +, \cdot, \langle \cdot, \cdot \rangle\big)$ is an* ***inner product space*** *iff the inner product operator $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ satisfies the following properties:*

(1) Commutativity: $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$ *for any* $\mathbf{u}, \mathbf{v} \in V$,

(2) Linearity: $\langle \lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2, \mathbf{v} \rangle = \lambda_1 \langle \mathbf{u}_1, \mathbf{v} \rangle + \lambda_2 \langle \mathbf{u}_2, \mathbf{v} \rangle$ *for any* $\lambda_1, \lambda_2 \in \mathbb{R}$ *and* $\mathbf{u}_1, \mathbf{u}_2, \mathbf{v} \in V$, *and*

(3) Positive definiteness: $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ *for any* $\mathbf{u} \in V$, *and equality holds iff* $\mathbf{u} = 0$.

Note that linearity also implies $\langle \mathbf{v}, \mathbf{0} \rangle = 0$ for any $\mathbf{v} \in V$, because

$$\langle \mathbf{v}, \mathbf{0} \rangle = \langle \mathbf{v}, 0\mathbf{u} \rangle = 0 \cdot \langle \mathbf{v}, \mathbf{u} \rangle = 0$$

where $\mathbf{u}$ is an arbitrary vector in V.

In an inner product space $\big(V, +, \cdot, \langle \cdot, \cdot \rangle\big)$, two vectors $\mathbf{v}$ and $\mathbf{u}$ are said to be **orthogonal** iff $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

A leading example of inner products is the **dot product** defined on $\mathbb{R}^n$. The dot product of two vectors $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$ is defined as

$$\mathbf{x} \cdot \mathbf{y} := \sum_{i=1}^{n} x_i y_i$$

Notice that the dot product above defined for two vectors is different from the scalar multiplication, which is defined for a scalar and a vector, although we usually use the same notation $\cdot$. It is straightforward to verify that the dot product satisfies our requirements on inner products.

An inner product induces a **norm** $\|\cdot\| : V \to \mathbb{R}_+$ by $\left[\|\mathbf{v}\| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}\right]$

Using the properties of inner product, it is straightforward to show that (1) $\|\mathbf{v}\| = 0$ iff $\mathbf{v} = \mathbf{0}$, and (2) $\|\lambda \mathbf{v}\| = |\lambda| \, \|\mathbf{v}\|$.

In $\mathbb{R}^n$, the norm induced by the dot product

$$\|\mathbf{x}\|_2 := \sqrt{\sum_{i=1}^{n} x_i^2}$$

is called the **Euclidean norm**, or L_2 **norm**.

Now let's look at an important inequality in inner product spaces.

Theorem 1.28 (Cauchy-Schwarz Inequality)**.** *In an inner product space* $\big(V, +, \cdot, \langle \cdot, \cdot \rangle\big)$, *we have*

$$\|\mathbf{u}\| \, \|\mathbf{v}\| \geq \big|\langle \mathbf{u}, \mathbf{v} \rangle\big|$$

for any $\mathbf{u}, \mathbf{v} \in V$.

Proof. If $\mathbf{u} = 0$, the inequality holds trivially. Now consider the case where $\mathbf{u} \neq \mathbf{0}$.

First, I claim that the vectors $\lambda \mathbf{u}$ and $\mathbf{v} - \lambda \mathbf{u}$ are orthogonal, where the real number λ is given by

$$\lambda := \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|^2}$$

This is because

$$\begin{aligned} \langle \lambda \mathbf{u}, \mathbf{v} - \lambda \mathbf{u} \rangle &= \lambda \langle \mathbf{u}, \mathbf{v} - \lambda \mathbf{u} \rangle = \lambda \big[\langle \mathbf{u}, \mathbf{v} \rangle - \lambda \langle \mathbf{u}, \mathbf{u} \rangle\big] \\ &= \lambda \left[\langle \mathbf{u}, \mathbf{v} \rangle - \lambda \|\mathbf{u}\|^2\right] = 0 \end{aligned}$$

Therefore, we have

$$\begin{aligned} \|\mathbf{v}\|^2 &= \big\|\lambda \mathbf{u} + (\mathbf{v} - \lambda \mathbf{u})\big\|^2 \\ &= \big\langle \lambda \mathbf{u} + (\mathbf{v} - \lambda \mathbf{u}), \lambda \mathbf{u} + (\mathbf{v} - \lambda \mathbf{u}) \big\rangle \\ &= \langle \lambda \mathbf{u}, \lambda \mathbf{u} \rangle + 2 \big\langle \lambda \mathbf{u}, (\mathbf{v} - \lambda \mathbf{u}) \big\rangle + \langle \mathbf{v} - \lambda \mathbf{u}, \mathbf{v} - \lambda \mathbf{u} \rangle \\ &= \lambda^2 \|\mathbf{u}\|^2 + \|\mathbf{v} - \lambda \mathbf{u}\|^2 \end{aligned}$$

As a result, we have $\|\mathbf{v}\|^2 \geq \lambda^2 \|\mathbf{u}\|^2$, i.e.

$$\|\mathbf{v}\|^2 \geq \left(\frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|^2}\right)^2 \|\mathbf{u}\|^2$$

i.e. $\|\mathbf{v}\|^2 \|\mathbf{u}\|^2 \geq \langle \mathbf{u}, \mathbf{v} \rangle^2$, and therefore we have $\|\mathbf{u}\| \, \|\mathbf{v}\| \geq \big|\langle \mathbf{u}, \mathbf{v} \rangle\big|$. □

Notice that Cauchy-Schwarz inequality also tells us that any norm induced by an inner product satisfies the **triangle inequality**: $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ for any $\mathbf{u}, \mathbf{v} \in V$. To see this,

$$\begin{aligned} \big(\|\mathbf{u}\| + \|\mathbf{v}\|\big)^2 &= \|\mathbf{u}\|^2 + 2 \|\mathbf{u}\| \, \|\mathbf{v}\| + \|\mathbf{v}\|^2 \geq \langle \mathbf{u}, \mathbf{u} \rangle + 2 \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\ &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle = \|\mathbf{u} + \mathbf{v}\|^2 \end{aligned}$$

and taking square root gives us $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

We can also take the triangle inequality as a part of the definition of the norm, and define the norm directly without the inner product.

Definition 1.29. *Let $(V,+,\cdot)$ be a vector space. The 4-tuple $\left(V,+,\cdot,\|\cdot\|\right)$ is a* ***normed vector space*** *iff the norm $\|\cdot\| : V \to \mathbb{R}_+$ satisfies the following properties:*

(1) $\|\mathbf{v}\| = 0$ iff $\mathbf{v} = \mathbf{0}$, for any $v \in V$,

(2) $\|\lambda\mathbf{v}\| = |\lambda|\,\|\mathbf{v}\|$, for any $\lambda \in \mathbb{R}$ and $\mathbf{v} \in V$, and

(3) Triangle inequality: $\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$, for any $\mathbf{u}, \mathbf{v} \in V$.

Clearly, if $\left(V,+,\cdot,\langle\cdot,\cdot\rangle\right)$ is an inner product space, and we define the norm $\|\cdot\| : V \to \mathbb{R}_+$ by $\|\mathbf{v}\| := \sqrt{\langle\mathbf{v},\mathbf{v}\rangle}$, then $\left(V,+,\cdot,\|\cdot\|\right)$ is a normed vector space. Especially, the triangle inequality is a corollary of Cauchy-Schwarz inequality. Therefore, normed vector spaces have less structures, and are more general than inner product spaces.

In a normed vector space, the norm also induces a metric by $d(\mathbf{u},\mathbf{v}) := \|\mathbf{u}-\mathbf{v}\|$. We can verify that this is indeed a metric on V. Clearly, $d(\mathbf{u},\mathbf{v}) = 0$ iff $\mathbf{u} = \mathbf{v}$ because of property (1) of the norm; $d(\mathbf{u},\mathbf{v}) = d(\mathbf{v},\mathbf{u})$ because of property (2) of the norm. Finally, a metric induced by the norm satisfies the triangle inequality because the norm satisfies the triangle inequality. To see this,

$$\begin{aligned} d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z}) &= \|\mathbf{x}-\mathbf{y}\| + \|\mathbf{y}-\mathbf{z}\| \geq \left\|(\mathbf{x}-\mathbf{y}) + (\mathbf{y}-\mathbf{z})\right\| \\ &= \|\mathbf{x}-\mathbf{z}\| = d(\mathbf{x},\mathbf{z}) \end{aligned}$$

Therefore, a normed vector space is automatically a metric space.

To summarize, in a vector space, an inner product induces a norm, which in turn induces a metric.

If we consider $\mathbb{R}^n$ endowed with the dot product as an inner space, then the dot product induces the L_2 norm, which in turn induces the Euclidean distance

$$d_2(\mathbf{x},\mathbf{y}) := \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Recall that we have shown that a valid inner product induces a valid norm, which in turn induces a valid metric. Because the dot product on $\mathbb{R}^n$ is a valid inner product, the Euclidean distance function d_2 is a valid metric; especially, it satisfies the triangle inequality required for metrics.

Note that for (1) a given vector space, there can exist more than one valid inner product; (2) for a given vector space, there can exist more than one valid norm, and a valid norm is not necessarily induced by some inner product, even if the space is an inner product space (for example the sup norm $||\mathbf{x}|| := \max_i |x_i|$ for $\forall\mathbf{x} \in \mathbb{R}^n$ in space $\mathbb{R}^n$ is not induced by any inner product); (3) for a given vector space, there can exist more than one valid metric, and a valid metric is not necessarily induced by some norm, even if the vector space is a normed vector space (for example the discrete metric is not induced by any norm).

2 Linear Functions on Vector Spaces

2.1 Definition, Examples

Definition 2.1. *Let E, F two vector spaces and $f : E \to F$ a function. We say that f is linear if :*

$$\begin{aligned} \forall(x,y) \in E^2, f(x +_E y) =& f(x) +_F f(y) \\ \forall x \in E, \forall\lambda \in \mathbb{R}, f(\lambda x) =& \lambda f(x) \end{aligned}$$

The two conditions can again be compacted into $f(\lambda x + y) = \lambda f(x) + f(y)$ for all x, y, λ. We denote by $\mathcal{L}(E, F)$ the set of linear functions from E to F. A linear applications from E to E is sometimes called an endomorphism of E; the set of such functions is denoted as $\mathcal{L}(E)$ (instead of $\mathcal{L}(E, E)$). If a linear function is bijective, it is called an isomorphism. A bibjective endomorphism is called an automorphism.

Example 2.2. *Let E a real vector space. Fix $\lambda \in \mathbb{R}$, and define the function :*

$$\begin{aligned} h : E &\to E \\ x &\mapsto \lambda x \end{aligned}$$

h is a linear function from E to E (an endomorphism).

Example 2.3. *Let $a, b, c \in \mathbb{R}$, the function :*

$$\begin{aligned} f : \mathbb{R}^3 &\to \mathbb{R} \\ (x, y, z) &\mapsto ax + by + cz \end{aligned}$$

is a linear function.

Example 2.4. *Let I an interval in $\mathbb{R}$. The derivation operator $\mathcal{D}$, understood as a function from the space of differentiable real functions over I, denoted $D(I, \mathbb{R})$ to the space of real function over I, is a linear function (you can verify that those two spaces are indeed vector spaces, although they are not finite dimensional). Indeed, we know :*

$$\mathcal{D}(\lambda f + g) = (\lambda f + g)' = \lambda f' + g' = \lambda \mathcal{D}(f) + \mathcal{D}(g)$$

2.2 Properties

Proposition 2.5. *Let $f : E \to F$ linear. Then :*

$$f(0_E) = 0_F$$
$$\forall x \in E, f(-x) = -f(x)$$

Proof.

$$f(0_E) = f(0 \cdot 0_E) = 0 \cdot f(0_E) = 0_F$$
$$f(-x) = f((-1) \cdot x) = (-1)f(x) = -f(x)$$

□

Theorem 2.6. *If $f : E \to F$ is linear, then :*

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

The proof of this theorem is left as an exercise.

In particular, $f(E)$ is a vector subspace of E called the Image or Range of f and denoted $Im(f)$; $f^{-1}(\{0_F\})$ is a vector subspace of E called the kernel of f and denoted $Ker(f)$. Those two sets provide a very neat characterization of injectivity and surjectivity for linear functions over vector spaces.

Proposition 2.7. *If $f : E \to F$ is linear, then :*

- *f is injective if and only if $Ker(f) = 0_E$*
- *f is surjective if and only if $Im(f) = E$*

Proof. The second statement is direct by definition. To prove the first equivalence, first assume f injective. Take $x \in Ker(f)$, i.e $f(x) = 0_F$. Since $f(0_E) = 0_F$, by injectivity $x = 0_E$, hence $Ker(f) = \{0_E\}$. For the other implication, assume $Ker(f) = \{0_E\}$. Consider x, x' such that $f(x) = f(x')$. By linearity $f(x - x') = 0_F$, hence $x - x' \in Ker(f)$ so $x - x' = 0_E$. In other words, $x = x'$. □

Lastly, we give some characterizations for the image of spanning and independent collections of vectors.

Proposition 2.8. *Let $f : E \to F$ a linear function. Then*

(i) If $(x_1, ..., x_n)$ a collection of vectors in E spans E, then $(f(x_1), ..., f(x_n))$ spans $Im(f)$

(ii) If $(x_1, ..., x_n)$ a collection of vectors in E is dependent, then $(f(x_1), ..., f(x_n))$ is dependent

(iii) f is injective if and only if for every independent collection of vectors in E $(x_1, ..., x_n)$, $(f(x_1), ..., f(x_n))$ is independent.

(iv) $(e_1, ..., e_n)$ is a basis for E and f is bijective if and only if $(f(x_1), ..., f(x_n))$ is a basis for F.

This highlights that linear functions are useful to rearrange vector spaces while preserving their properties. A consequence of the previous proposition is that we can express a change of coordinates as a linear function. It turns out that this is actually an equivalence. The usefulness of linear functions between vector spaces leads us to study another object that can actually serve as an equivalent representation : matrices.

Before we move to the next section, we just state without proof one important theorem for linear functions between finite dimensional vector spaces.

Theorem 2.9. *Let $f : E \to F$ a linear function. We have :*

$$dim(E) = dim(Ker(f)) + dim(Im(f))$$

Furthermore :

$$dim(Im(f)) \leq \min(dim(E), dim(F))$$

and f is injective if and only if $dim(Im(f)) = dim(E)$, f is surjective if and only if $dim(Im(f)) = dim(F)$.

3 Matrices

3.1 Definition

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^{th}**-entry** or the ij^{th} **-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

3.2.1 Addition and scalar multiplication

Along with the addition and scalar multiplications operations, the set of matrices $\mathcal{M}_{mn}$ is going to be a vector space. The two operations 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 $\lambda \in \mathbb{R}$.*
(1) The sum $A + B$ is the matrix whose ij-entry is $a_{ij} + b_{ij}$.
(2) The scalar multiplication of A by λ, λA is the matrix whose ij^{th}-entry is λa_{ij}.

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

Proposition 3.3. *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^{th} entry which is equal to 1, then $(E_{ij})_{i=1...m}^{j=1...n}$ is a basis of $\mathcal{M}_{mn}$. It is called the canonical basis of $\mathcal{M}_{mn}$.

3.2.2 Multiplication

The matrix 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 3.4. *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^{th}-entry is:*

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

In fact, matrix multiplication should be interpreted as a linear mapping. Recall that in an $m \times n$ matrix with components in $\mathbb{R}$, each column can be viewed as a vector in $\mathbb{R}^m$, and each row can be viewed as a vector in $\mathbb{R}^n$. If we left-multiply matrix A by another matrix C, each row of the product CA is a linear combination of the rows of A; therefore, left-multiplying a matrix can be interpreted as a row transformation. If we right-multiply A by another matrix C, each column of the product AC is a linear combination of the columns of A; therefore, right-multiplying a matrix should be interpreted as a column transformation. The following properties of matrix multiplication are easy to verify from the definition.

Proposition 3.5. *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. addition: $A(B + C) = AB + AC$ and $(B + C)A = BA + CA$.*
- *The multiplication is associative: $A(BC) = (AB)C$.*[2]
- *$A(\lambda B) = \lambda(AB)$.*

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.

[2] To see this, first, $(AB)\,C$ and $A\,(BC)$ have the same size $m \times q$. Second, for any $(i, l) \in \{1, \ldots, m\} \times \{1, \ldots, q\}$, we have

$$\begin{aligned}
\left((AB)\,C\right)_{il} &= \sum_{k=1}^{p} (AB)_{ik}\, c_{kl} = \sum_{k=1}^{p} \left[\left(\sum_{j=1}^{n} a_{ij} b_{jk} \right) c_{kl} \right] = \sum_{k=1}^{p} \left(\sum_{j=1}^{n} a_{ij} b_{jk} c_{kl} \right) \\
&= \sum_{j=1}^{n} \left(\sum_{k=1}^{p} a_{ij} b_{jk} c_{kl} \right) = \sum_{j=1}^{n} \left(a_{ij} \sum_{k=1}^{p} b_{jk} c_{kl} \right) = \sum_{j=1}^{n} \left(a_{ij}\, (BC)_{jl} \right) \\
&= \left(A\,(BC) \right)_{il}.
\end{aligned}$$

Definition 3.6. *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 B, so the inverse of the inverse of A is A itself: $(A^{-1})^{-1} = A$. Besides:

Proposition 3.7. *If $A, B \in \mathcal{M}_{nn}$ are invertible, then so is their product AB and:*

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

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

Definition 3.8. *$A^k = A...A$ 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 3.9. *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): $A^T = (a_{ji})$.*

Obviously, if we apply the transpose operator twice, we end up back on A: $(A^T)^T = A$. 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 3.10.

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

Definition 3.11. *A square matrix A is* ***symmetric*** *iff it is equal to its transpose $A' = A$.*

3.2.4 Rank

If A is an $m \times n$ real 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 3.12. *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 3.13. *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 3.14. *The row rank of a matrix is equal to its column rank.*

Because the rows of the product matrix AB are linear combinations of rows of B, the basis of the rows of B can represent rows of AB, and so we have $Rank\,(AB) \leq Rank\,(B)$. Because the columns of AB are linear combinations of columns of A, the basis of columns of A can represent columns of AB, and so we have $Rank\,(AB) \leq Rank\,(A)$. As a result, we always have

$$Rank\,(AB) \leq \min\left\{Rank\,(A)\,, Rank\,(B)\right\}$$

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

Proposition 3.15. *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 $BA = C'ABA = (C'A)(BA) = C'(AB)A = C'A = I_n$.

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}, $\lambda = A^{-1}0 = 0$. $\square$

3.2.5 Trace

Definition 3.16. *Let $A = (a_{ij})$ be a square matrix of size n. The* ***trace*** *of A, noted $tr(A)$, is* $tr(A) = \displaystyle\sum_{i=1}^{n} a_{ii}$.

The trace has the following properties:

Proposition 3.17.

- *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), $tr(AB) = tr(BA)$.*

3.2.6 Determinants

Definition 3.18. *For a square matrix A, its* ***determinant****, denoted as* $\det(A)$*, is an element defined inductively in the following way:*

(1) For a 1×1 matrix $A = a_{11}$, define its determinant as $\det(A) := a_{11}$*.*

(2) For an $n \times n$ matrix where $n \geq 2$, define its determinant as

$$\det(A) := \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det\left(A_{-1,-j}\right)$$

where $A_{-i,-j}$ is the matrix A with the i-th row and j-th column eliminated.

According to the definition above, for a 2×2 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

its determinant is $\det A = a_{11}a_{22} - a_{12}a_{21}$.

For a 3×3 matrix

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

its determinant is

$$\begin{aligned} \det(A) &= a_{11} \det\left(\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}\right) - a_{12} \det \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} + a_{13} \det \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \end{aligned}$$

For a 4×4 matrix A, we should expect that $\det(A)$ has $4! = 24$ terms, and so we don't bother to write it down here.

In the inductive definition of determinants above, the induction formula

$$\det(A) := \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det\left(A_{-1,-j}\right)$$

is also called the **cofactor expansion** of A along the first row. The (i,j)-th **cofactor** of a square matrix A, denoted as A_{ij}, is defined as

$$A_{ij} := (-1)^{i+j} \det\left(A_{-i,-j}\right)$$

and so the cofactor expansion of A along the first row can be rewritten as

$$\det(A) := \sum_{j=1}^{n} a_{1j} A_{1j}$$

In fact, we can equivalently define determinants by expanding along any row or column, i.e.

$$\det(A) := \sum_{j=1}^{n} a_{ij} A_{ij}$$

for any arbitrary row i, or

$$\det(A) := \sum_{i=1}^{n} a_{ij} A_{ij}$$

for any arbitrary column j. Let's admit the equivalence of different ways of expansion without providing a proof.

Using this equivalence of expanding along a row or a column, we can show that $\det\left(A^T\right) = \det(A)$.

We also state the following two results without proof.

Theorem 3.19. *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. *The determinant of a triangular (this includes diagonal) matrix $A = (a_{ij})$ is the product of its diagonal elements* $\det(A) = \prod_{i=1}^{n} a_{ii}$.
3. ***Multilinearity**: $\det$ is linear with respect to each of its argument:*

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

4. *If any two columns of A are equal, then* $\det(A) = 0$.
5. ***Antisymmetry**: If two columns of A are interchanged, then the determinant changes by a sign.*
6. *If one adds a scalar multiple of one column to another then the determinant does not change.*

Theorem 3.20. *(1) A matrix and its transpose have the same determinant:* $\det(A') = \det(A)$. *(Equivalently, n vectors $A^1, ..., A^n$ of $\mathbb{R}^n$ are linearly independent iff* $\det(A^1, ..., A^n) \neq 0$.*)*

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

(3) For two $n \times n$ matrices, we have $\det(AB) = \det(A)\det(B)$.

Clearly, the theorem above implies that $\det\left(A^{-1}\right) = \left(\det(A)\right)^{-1}$ for an invertible matrix A.

4 Systems of Linear Equations

Consider the following system of linear equations in x:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

where a_{ij}'s and b_i's are all elements of $\mathbb{R}$, and the unknowns $x_1, \ldots, x_n$ also take values in $\mathbb{R}$.

We can rewrite the system of linear equations in a compact way by

$$A\mathbf{x} = \mathbf{b}$$

where $A = \left(a_{ij}\right)$ is the $m \times n$ matrix, $\mathbf{b} = (b_1, \ldots, b_m)^T$, and $\mathbf{x} = (x_1, \ldots, x_n)^T$.

If we view $\mathbf{x}$ as a column transformation of the columns of A, then the equation asks us to find ways to represent the vector $\mathbf{b} \in \mathbb{R}^m$ as a linear combination of the columns of A. To clearly see this, write matrix A as $A = (\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n)$, where $\mathbf{a}_i$ is the ith column of A, we have $(\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n)\mathbf{x} = (\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n)\left(x_1, \ldots, x_n\right)^T = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + ... + x_n\mathbf{a}_n$.

Proposition 4.1. *The system of equations $A\mathbf{x} = \mathbf{b}$ (A is an $m \times n$ matrix over $\mathbb{R}$) has a solution iff*

$Rank\left(\left[A|\mathbf{b}\right]\right) = Rank\,(A)$.

When the system has a solution,

(1) the solution is unique iff the columns of A are linearly independent, i.e. $Rank\,(A) = n$.

(2) the system has infinitely many solutions iff $Rank\,(A) < n$.

To see this, suppose $Rank\,(A) = n$, then the columns of A are linearly independent, and therefore their linear representation of b is unique. On the other hand, suppose $Rank\,(A) < n$, then the columns of A are linearly dependent. Therefore, they have a non-trivial linear representation of the zero vector, i.e. there exists $F \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ s.t. $A\mathbf{z} = \mathbf{0}$. Then if $\mathbf{x}^*$ is a solution to the system, then $\mathbf{x}^* + \lambda\mathbf{z}$ is also a solution, for any $\lambda \in \mathbb{R}$, and so the solution is not unique.

Proposition 4.2 (General solution of a linear equation)**.** *Let a vector $\mathbf{x}^*$ satisfy the equation $A\mathbf{x}^* = \mathbf{b}$, and let $H := \{\mathbf{z} : A\mathbf{z} = \mathbf{0}\}$. Then the set $\{\mathbf{x} = \mathbf{x}^* + \mathbf{x}_h : \mathbf{x}_h \in H\}$ is the set of all solutions of the equation $A\mathbf{x} = \mathbf{b}$.*

The system of linear equations can be solved by hand using Gauss-Jordan elimination, which is essentially row operations of the matrix $\left[A|b\right]$. You may refer to a standard linear algebra textbook for details.

As a special case, when $m = n$ and the square matrix A is invertible, the unique solution is clearly $\mathbf{x}^* = A^{-1}b$. We also have an explicit formula for $x^* = A^{-1}b$, which is known as Cramer's rule. Let's state it without proof.

Theorem 4.3 (Cramer's Rule)**.** *Let A be an $n \times n$ invertible matrix and b be an $n \times 1$ column vector. The i-th entry of the $n \times 1$ column vector $\mathbf{x}^* := A^{-1}\mathbf{b}$ can be calculated as*

$$x_i^* = \frac{\det\left(A_i\right)}{\det\left(A\right)}$$

for each i, where A_i is the $n \times n$ matrix formed by replacing the i-th column of A by $\mathbf{b}$ and leaving the other columns unchanged.

However, calculating determinants is numerically difficult when the size of the matrices is large, since the determinant of an $n \times n$ matrix has $n!$ terms. So Cramer's rule may not be as useful as it seems.

5 Eigenvalues, Eigenvectors, and Diagonalization

5.1 Eigenvalues and Eigenvectors

The concept of eigenvalues is especially important in linear dynamic systems.

Definition 5.1. *Let A be an $n \times n$ matrix over $\mathbb{C}$. A scalar $\lambda \in \mathbb{C}$ is said to be an* ***eigenvalue of*** *A iff $\exists\ \mathbf{x} \in \mathbb{C}^n \backslash \{\mathbf{0}\}$ s.t. $A\mathbf{x} = \lambda\mathbf{x}$. A vector $\mathbf{x} \in \mathbb{C}^n \backslash \{\mathbf{0}\}$ is said to be an* ***eigenvector of*** *A iff $\exists\ \lambda \in \mathbb{C}$ s.t. $A\mathbf{x} = \lambda\mathbf{x}$.*

Proposition 5.2. $\lambda \in \mathbb{C}$ *is an eigenvalue of* A *iff* $\det(\lambda I_n - A) = 0$.

By definition, $\lambda \in \mathbb{C}$ is an eigenvalue of A iff $A\mathbf{x} = \lambda\mathbf{x}$ has a nonzero solution. This is equivalent to $(A - \lambda I_n)\mathbf{x} = 0$ having a nonzero solution, which is in turn equivalent to the columns of the matrix $\lambda I_n - A$ being linearly dependent, which is in turn equivalent to $\det(\lambda I_n - A) = 0$.

In the determinant of the matrix

$$\lambda I_n - A = \begin{bmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{bmatrix}$$

the diagonal contributes a term λ^n, and all other terms have a degree no more than $n-2$. Therefore, $\det(\lambda I_n - A)$ is a polynomial of λ of degree n. The polynomial $P_A(\lambda) := \det(\lambda I_n - A)$ is also called the **characteristic polynomial of** A. By construction, $\lambda \in \mathbb{C}$ is an eigenvalue of A iff $P_A(\lambda) = 0$.

Theorem 5.3 (Fundamental Theorem of Algebra)**.** *Let* $P : \mathbb{C} \to \mathbb{C}$ *be a polynomial of degree* n*, i.e.* $P(\lambda) = c_n\lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0$*, where* $c_k \in \mathbb{C}$ *for any* $k = 0, 1, \ldots, n$ *and* $c_n \neq 0$*. Then* P *has exactly* n *roots in* $\mathbb{C}$*, counted with multiplicity. That is, there exists* $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{C}$ *s.t.*

$$P(\lambda) = c_n(\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$$

Therefore, we can obtain all eigenvalues of A by setting the characteristic polynomial of A to 0 and solving for all its roots.

Proposition 5.4. *Let* A *be an* $n \times n$ *matrix over* $\mathbb{C}$ *and* $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{C}$ *are eigenvalues of* A*. The the characteristic function polynomial of* A*:*

$P_A(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)...(\lambda - \lambda_n)$.

Corollary 5.5. *Let* A *be an* $n \times n$ *matrix over* $\mathbb{C}$ *and* $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{C}$ *are eigenvalues of* A*. Then* $\det(A) = \lambda_1\lambda_2...\lambda_n$.

Proof. Plug $\lambda = 0$ in $P_A(\lambda) = \det(\lambda I_n - A)$, we have $\det(-A) = (-1)^n \det(A) = (0 - \lambda_1)(0 - \lambda_2)...(0 - \lambda_n) = (-1)^n\lambda_1\lambda_2...\lambda_n$. □

Suppose we have established that $\lambda \in \mathbb{C}$ is an eigenvalue of A. Then we can obtain the set of all eigenvectors associated with λ by solving for all nonzero solutions in $\mathbb{C}^n$ to the system of linear equations $(A - \lambda I_n)\mathbf{x} = 0$.

5.2 Diagonalization

Definition 5.6. *Let* A *be an* $n \times n$ *matrix over* $\mathbb{C}$*. The matrix* A *is* ***diagonalizable in*** $\mathbb{C}$ *iff there exists an* $n \times n$ *invertible matrix* P *over* $\mathbb{C}$ *and an* $n \times n$ *diagonal matrix* Λ *over* $\mathbb{C}$ *s.t.* $P^{-1}AP = \Lambda$*. The matrix* A *is* ***diagonalizable in*** $\mathbb{R}$ *iff there exists an* $n \times n$ *invertible real matrix* P *and an* $n \times n$ *diagonal real matrix* Λ *s.t.* $P^{-1}AP = \Lambda$*.*

Intuitively, the matrix A is diagonalizable iff we can find invertible matrix P s.t. we can transform A into some diagonal matrix by left-multiplying A by P^{-1} and right-multiplying A by P.

If A can be diagonalized as Λ using P, then

$$\begin{aligned} \det(\lambda I_n - A) &= \det\left(P^{-1}\right)\det(\lambda I_n - A)\det(P) = \det\left[P^{-1}(\lambda I_n - A)P\right] \\ &= \det\left(\lambda I_n - P^{-1}AP\right) = \det(\lambda I_n - \Lambda) \\ &= (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n) \end{aligned}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the n entries on the diagonal of the matrix Λ. Therefore, the entries on the diagonal of Λ must be the n eigenvalues of A.

When a matrix A is diagonalizable, i.e. $P^{-1}AP = \Lambda$, we have

$$A = \left(PP^{-1}\right) A \left(PP^{-1}\right) = P\left(P^{-1}AP\right)P^{-1} = P\Lambda P^{-1}$$

and so the matrix A can be decomposed as $P\Lambda P^{-1}$.

Not all matrices are diagonalizable. For example, consider the matrix

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

It is straightforward to see that the two eigenvalues of A are both 0. So if A can be diagonalized as Λ under P, i.e. $P^{-1}AP = \Lambda$, the diagonal matrix Λ must be the 2×2 zero matrix. Then we have $A = P\Lambda P^{-1} = 0$. Contradiction.

The next proposition establishes a necessary and sufficient characterization of diagonalizable matrices.

Proposition 5.7. *Let A be an $n \times n$ matrix over $\mathbb{C}$. Then A is diagonalizable in $\mathbb{C}$ iff A has n linearly independent eigenvectors.*

Proof. $\Leftarrow$:

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in \mathbb{C}^n \setminus \{\mathbf{0}\}$ be the n linearly independent eigenvectors of A, and let the corresponding eigenvalues be $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{C}$. By definition, we have $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$ for each i. Therefore,

$$\begin{aligned} A\left[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\right] &= \left[A\mathbf{x}_1, A\mathbf{x}_2, \ldots, A\mathbf{x}_n\right] \\ &= \left[\lambda_1\mathbf{x}_1, \lambda_2\mathbf{x}_2, \ldots, \lambda_n\mathbf{x}_n\right] \\ &= \left[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\right] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \end{aligned}$$

Let $P := [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$ and

$$\Lambda := \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

we have $AP = P\Lambda$. Because $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are linearly independent, the matrix P is invertible. Therefore, we have $P^{-1}AP = P^{-1}P\Lambda = \Lambda$, and so A is diagonalizable.

$\Rightarrow$:

Because A is diagonalizable, there exists invertible P and diagonal Λ s.t. $P^{-1}AP = \Lambda$. Rewrite the equality as $AP = P\Lambda$, i.e.

$$A\left[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\right] = \left[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\right] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

where $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are columns of P and $\lambda_1, \lambda_2, \ldots, \lambda_n$ are entries on the diagonal of Λ. Then for each i, we have $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$. Because P is invertible, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are not the zero vector, and therefore $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are eigenvectors. Again by invertibility of P, we know that $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are linearly independent. $\square$

The next proposition states that if an $n \times n$ matrix A has n distinct eigenvalues, then it has n linearly independent eigenvectors.

Proposition 5.8. *Let A be an $n \times n$ matrix over $\mathbb{C}$ with n distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{C}$. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in \mathbb{C}^n \backslash \{\mathbf{0}\}$ be the corresponding eigenvectors. Then $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are linearly independent.*

Let's show a weaker version of the proposition: the eigenvectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are linearly independent if the corresponding λ_1 and λ_2 are distinct. To see this, suppose $\mathbf{x}_1$ and $\mathbf{x}_2$ are linearly dependent, i.e. there exists $k_1, k_2 \in \mathbb{C}$ s.t. at least one is not 0 and $k_1 x_1 + k_2 x_2 = 0$. Clearly, because $x_1 \neq 0$ and $x_2 \neq 0$, we have $k_1 \neq 0$ and $k_2 \neq 0$, since if one of k_i is 0 then the other is also 0. Therefore,

$$A\mathbf{x}_1 = \lambda_1 \mathbf{x}_1 = \lambda_1 \left(-k_2/k_1 \cdot \mathbf{x}_2\right) = -k_2 \lambda_1 / k_1 \cdot \mathbf{x}_2$$

and

$$A\mathbf{x}_1 = A\left(-k_2/k_1 \cdot \mathbf{x}_2\right) = -k_2/k_1 \cdot A\mathbf{x}_2 = -k_2 \lambda_2 / k_1 \cdot \mathbf{x}_2$$

Comparing the two equations above, we have $(\lambda_1 - \lambda_2)\, k_2/k_1 \cdot \mathbf{x}_2 = 0$. Because $\mathbf{x}_2 \neq 0$ and $k_2 \neq 0$, we have $\lambda_1 = \lambda_2$, which contradicts the assumption that λ_1 and λ_2 are distinct.

The argument above only proves the proposition when $n = 2$, but we should expect the proof for the general statement to be similar. Let's skip the proof for the general statement and admit the result.

Combining the two propositions above, we have the following theorem.

Theorem 5.9. *Let A be an $n \times n$ matrix over $\mathbb{C}$ with n distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{C}$. Then A is diagonalizable in $\mathbb{C}$.*

If the eigenvalues of A are not distinct, we don't know whether A is diagonalizable or not.

A matrix over field $\mathbb{R}$ is called a **real matrix**. An $n \times n$ matrix A is **symmetric** iff $a_{ij} = a_{ji}$ for any i and j. An $n \times n$ matrix A over F is **orthogonal** iff $A^T A = I_n$. Clearly, the condition $A^T A = I_n$ means that the columns of A are pairwise orthogonal (w.r.t. dot product) and each have a norm of 1. The condition $A^T A = I_n$ also implies that A is invertible and $A^{-1} = A^T$, and the columns of A are a basis of F^n.

The next theorem states that a real symmetric matrix is always diagonalizable in $\mathbb{R}$, i.e. there exists a real invertible matrix P and a real diagonal matrix Λ s.t. $P^{-1} A P = \Lambda$. This result is a little involved, and let's state it without proof.

Theorem 5.10. *Let A be an $n \times n$ real symmetric matrix. Then all its eigenvalues are real, and there exists a real orthogonal matrix P and a real diagonal matrix Λ s.t. $P^{-1} A P = P^T A P = \Lambda$.*

For an economist, the motivation for studying eigenvalues, eigenvectors, and diagonalization is their applications in dynamic models.

Consider a linear dynamic system $\mathbf{x}_t = A\mathbf{x}_{t-1}$, where $\mathbf{x}_t$ is an n-dimensional real vector and A is an $n \times n$ real matrix. Clearly we have $\mathbf{x}_t = A^t \mathbf{x}_0$. When t is large, it is difficult to analyze the behavior of x_t since calculating A^t is difficult. With the help of diagonalization, however, $A^t = \left(P \Lambda P^{-1}\right)^t = P \Lambda^t P^{-1}$, where Λ^t is easy to calculate since Λ is diagonal. In fact, if all eigenvalues of A have a modulus strictly less than 1, then $\Lambda^t \to 0$ (the $n \times n$ zero matrix), and so $\mathbf{x}_t = P\Lambda^t P^{-1}\mathbf{x}_0 \to \mathbf{0}$ (the n-dimensional zero vector).

If the dynamic system is not linear, it is a standard practice in macro to log-linearize the dynamic system around its steady state, which is essentially approximating a non-linear system using a linear system. Then our discussion on linear dynamic systems above applies.

6 Quadratic Forms

Definition 6.1. *A quadratic form on $\mathbb{R}^n$ is a function $Q : \mathbb{R}^n \to \mathbb{R}$ that can be represented by*

$$\begin{aligned} Q(\mathbf{x}) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j \\ &= a_{11}x_1^2 + a_{22}x_2^2 + \cdots + a_{nn}x_n^2 \\ &+ (a_{12} + a_{21})\, x_1x_2 + (a_{13} + a_{31})\, x_1x_3 \\ &+ \cdots + \left(a_{n-1,n} + a_{n,n-1}\right) x_{n-1}x_n \end{aligned}$$

where a_{ij}'s are real coefficients.

Notice that we can write a quadratic form in a compact way using matrix multiplication. The quadratic form $Q(\mathbf{x})$ defined above is equal to $\mathbf{x}^T A\mathbf{x}$, where $A = \left(a_{ij}\right)$ is the $n \times n$ matrix whose elements are the coefficients of the quadratic form, and $\mathbf{x}$ is considered as a column vector.

The way to represent a quadratic form Q using a matrix A is not unique, since if the matrix A represents the quadratic form Q, then the matrix $A + B$ also represents Q for any *anti-symmetric* matrix B (i.e. $b_{ij} = -b_{ji}$ for any i, j). However, each quadratic form Q can be represented by a unique symmetric matrix A.

Definition 6.2. *Let A be an $n \times n$ real symmetric matrix. The matrix A, or the quadratic form $Q(x) := \mathbf{x}^T A\mathbf{x}$ that is represented by A, is said to be*

(1) ***positive definite****, iff $\mathbf{x}^T A\mathbf{x} > 0$ for any $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$;*
(2) ***negative definite****, iff $\mathbf{x}^T A\mathbf{x} < 0$ for any $x \in \mathbb{R}^n \setminus \{\mathbf{0}\}$;*
(3) ***positive semi-definite****, iff $\mathbf{x}^T A\mathbf{x} \geq 0$ for any $x \in \mathbb{R}^n$;*
(4) ***negative semi-definite****, iff $\mathbf{x}^T A\mathbf{x} \leq 0$ for any $x \in \mathbb{R}^n$;*
(5) ***indefinite****, iff $\exists\ \mathbf{x}, \mathbf{x}' \in \mathbb{R}^n$ s.t. $\mathbf{x}^T A\mathbf{x} > 0$ and $\mathbf{x}'^T A\mathbf{x}' < 0$.*

The next theorem provides a necessary and sufficient characterization of positive/negative (semi-)definite matrices using eigenvalues.

Theorem 6.3. *Let A be an $n \times n$ real symmetric matrix. The matrix A is*

(1) positive definite iff all its eigenvalues are positive;
(2) negative definite iff all its eigenvalues are negative;
(3) positive semi-definite iff all its eigenvalues are non-negative;
(4) negative semi-definite iff all its eigenvalues are non-positive;
(5) indefinite iff it has both positive and negative eigenvalues.

This theorem easily follows Theorem 5.10, which allows us to find orthogonal P and diagonal Λ s.t. $P^T AP = \Lambda$, since we have

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T\left(P\Lambda P^T\right)\mathbf{x} = \left(P^T\mathbf{x}\right)^T \Lambda \left(P^T\mathbf{x}\right) = \mathbf{y}^T\Lambda\mathbf{y} = \sum_{i=1}^{n}\lambda_i y_i^2$$

where $\mathbf{y} := P^T\mathbf{x}$, and λ_i's are entries on the diagonal of Λ.

There are some other characterizations of positive/negative (semi-)definiteness using *principal minors*. Please refer to FMEA Theorem 1.7.1 and 1.8.1 for details.

Theorem 6.4 (LDL Decomposition)**.** *Let A be an $n \times n$ real symmetric matrix. Then A is positive definite iff there exists a real diagonal matrix D with positive entries on its diagonal and a real lower triangle matrix*[3] *L with all 1's on its diagonal, s.t. $A = LDL^T$.*

The decomposition $A = LDL^T$ in the theorem is called the **LDL decomposition** of a positive definite matrix. The "if" part of the theorem is straightforward, but the "only if" part is involved. Let's admit this result without proof.

In the theorem above, if we define $P := L\sqrt{D}$, where $\sqrt{D}$ is the diagonal matrix whose entries on its diagonal are the square root of the corresponding entries of D, then we have $A = PP^T$. This is called the **Choleski decomposition** of the positive definite matrix A.

Theorem 6.5 (Cholesky Decomposition)**.** *Let A be an $n \times n$ real symmetric matrix. Then A is positive definite iff there exists a real lower triangle matrix P with all positive entries on its diagonal s.t. $A = PP^T$.*

[3] An $n \times n$ matrix is said to be a **lower triangle matrix** iff $a_{ij} = 0$ for any $i < j$.