Linear Algebra

Daniel Loghin

School of Mathematics

$$\begin{bmatrix} \boldsymbol{\times} & \times & \times \\ \boldsymbol{\times} & \times & \times \\ \boldsymbol{\times} & \times & \times \\ \boldsymbol{\times} & \times & \times \end{bmatrix} = \begin{bmatrix} + & \times & \times \\ + & \times & \times \\ + & \times & \times \\ + & \times & \times \end{bmatrix} \begin{bmatrix} \oplus & & \\ & 1 & \\ & & 1 \end{bmatrix}$$

$$\begin{bmatrix} \times & \boldsymbol{\times} & \times \\ \times & \boldsymbol{\times} & \times \\ \times & \boldsymbol{\times} & \times \\ \times & \boldsymbol{\times} & \times \end{bmatrix} = \begin{bmatrix} + & + & \times \\ + & + & \times \\ + & + & \times \\ + & + & \times \end{bmatrix} \begin{bmatrix} + & \oplus & \\ & \oplus & \\ & & 1 \end{bmatrix}$$

$$\begin{bmatrix} \times & \times & \boldsymbol{\times} \\ \times & \times & \boldsymbol{\times} \\ \times & \times & \boldsymbol{\times} \\ \times & \times & \boldsymbol{\times} \end{bmatrix} = \begin{bmatrix} + & + & + \\ + & + & + \\ + & + & + \\ + & + & + \end{bmatrix} \begin{bmatrix} + & + & \oplus \\ & + & \oplus \\ & & \oplus \end{bmatrix}$$

Lecture Notes: 2LA/2LALP(3): Autumn 2024

Contents

I. Vector spaces

LECTURE 1

Background

The definition of vectors varies across disciplines and textbooks.

In science and engineering, vectors are concepts used for modeling various physical phenomena. Examples of physical concepts modeled as vectors include position, velocity, acceleration, force, momentum etc. Since many physical quantities depend on time or position, vectors often turn out to be (or are used to mean) vector functions.

In mathematics, vectors represent abstract objects which satisfy certain axioms and form a so-called vector space. They are studied independently of their physical meaning/motivation.

In this lecture, we look at the motivation behind the concept of vector space and justify the abstract setting that will be assumed throughout the course.

1.1 Motivation

The concept of a vector arose relatively late in the 19th century[1], despite the evident need in geometry and mechanics. We start with a brief discussion of 3D vectors; this will allow us to extract the axioms used to formulate the concept of a vector space.

1.1.1 Vectors in mechanics and geometry

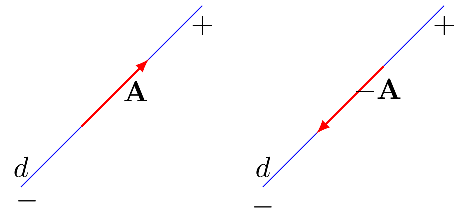
> **Definition 1.1** Vectors are modeling objects equipped with **magnitude**, **direction line** d and **orientation** (positive or negative) relative to d.

We denote vectors by bold uppercase or lowercase letters, e.g., $\mathbf{A}$, with magnitude indicated by $|\mathbf{A}|$, or $\mathbf{v}$ with magnitude indicated by $\|\mathbf{v}\|$. The familiar 3D space will be referred to as **Euclidean space** and will be denoted by $\mathbb{E}^3$. For now, we can view $\mathbb{E}^3$ as the set of 3D vectors.

A vector is represented geometrically in $\mathbb{E}^3$ by an arrow. Given a (pre-assigned) positive orientation

[1]Encyclopedia Brittanica: vectors appeared late in the 19th century when Josiah Willard Gibbs and Oliver Heaviside (of the United States and Britain, respectively) independently developed vector analysis to express the new laws of electromagnetism discovered by the Scottish physicist James Clerk Maxwell.

relative to a direction line, a vector is denoted by $+\mathbf{A}$, or $\mathbf{A}$, if the arrow points in the positive direction of d, and by $-\mathbf{A}$ if the arrow points in the negative directon of d.

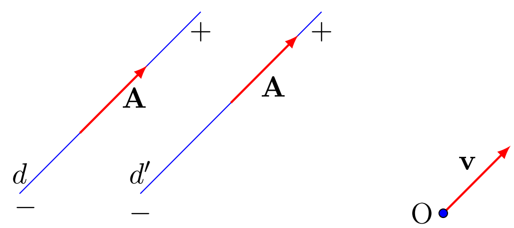


Vectors $\mathbf{A}$ and $-\mathbf{A}$.

We distinguish three types of vectors:

1. **Free vectors** These are hop-and-slide vectors: they are not bound to an origin, nor to a direction line, but are allowed to 'hop' on to any other line parallel to a given direction line and 'slide' up or down on it, while preserving their orientation and magnitude.
2. **Sliding vectors** These vectors are allowed to slide on a given (fixed) direction line, but cannot 'hop' onto any other line.
3. **Bound vectors** are vectors equipped additionally with an origin, namely the tail of a bound vector is fixed at a given point (the origin).

R There is an exception to the above definition and classification: **the zero vector**, which is the unique vector with no direction and zero magnitude.



Free vector $\mathbf{A}$ Bound vector $\mathbf{v}$

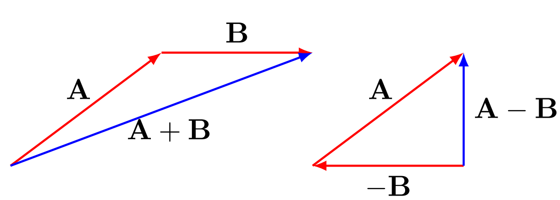
Free vectors are not uniquely defined geometrically: any two oriented segments of equal length with parallel direction lines represent the same vector. Bound vectors are uniquely defined, given a fixed origin. In applications, both types arise.

1.1.2 Operations with vectors

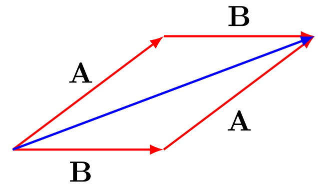
All operations with vectors arise from practical considerations and observe the axioms and results of geometry. This is evident for the first two operations.

Addition and scalar multiplication

Definition 1.2 — Vector addition. The sum of two vectors $\mathbf{A}, \mathbf{B}$ is defined via the head-to-tail procedure, also known as the parallelogram law: the tail of $\mathbf{B}$ is placed at the head of $\mathbf{A}$ and the sum is defined as the vector represented by the segment from the tail of $\mathbf{A}$ to the head of $\mathbf{B}$. The operation of **subtraction of two vectors** is defined as the addition of vectors $\mathbf{A}$ and $-\mathbf{B}$.



Using Definition 1.2, we can use the concept of free vectors to form the parallelogram below and thus observe that the operation $\mathbf{B}+\mathbf{A}$ yields the same vector (blue segment in figure below) as the operation $\mathbf{A}+\mathbf{B}$, i.e., $\mathbf{B}+\mathbf{A}=\mathbf{A}+\mathbf{B}$. Thus, vector addition as defined above is a **commutative operation**.



Exercise 1.1 Use a geometrical argument to show that the sum of three vectors is the same irrespective of their order in the sum.

The above exercise shows that vector addition as outlined in Definition 1.2 is **associative**.

Before we define the next operation, we note that, geometrically, the vector $\mathbf{A}+\mathbf{A}$ has magnitude $2|\mathbf{A}|$. This justifies us to define the vector $2\cdot\mathbf{A}$, i.e., 'two copies of $\mathbf{A}$', to represent the sum $\mathbf{A}+\mathbf{A}$. This observation is generalised in the following definition.

Definition 1.3 — Scalar-vector product. Let $c\in\mathbb{R}$. The scalar-vector product $c\cdot\mathbf{A}$, or simply $c\mathbf{A}$, is a vector with magnitude $|c|\,|\mathbf{A}|$ and same direction as $\mathbf{A}$ if c is positive and opposite direction if c is negative.

Exercise 1.2 Let $a,b\in\mathbb{R}$. Use Definition 1.3 to derive the following properties:

i. $(a+b)\mathbf{A}=a\mathbf{A}+b\mathbf{A}$;
ii. $a(\mathbf{A}+\mathbf{B})=a\mathbf{A}+a\mathbf{B}$.

The above exercise shows that the scalar-vector product is a **distributive operation**.

This definition allows to introduce the useful concept of a **unit vector**. Thus, we can associate with any vector $\mathbf{A}$ a vector of unit length, also known as **unit vector**, with the same orientation as $\mathbf{A}$. This is done using the scalar-vector product by multiplying $\mathbf{A}$ by $|\mathbf{A}|^{-1}$:

$$\widehat{\mathbf{A}}:=|\mathbf{A}|^{-1}\,\mathbf{A}=\frac{1}{|\mathbf{A}|}\mathbf{A}=\frac{\mathbf{A}}{|\mathbf{A}|}.$$

Dot product

Let $\mathbf{A},\mathbf{B}$ be vectors in $\mathbb{E}^3$; let α be the angle between the directions of $\mathbf{A}$ and $\mathbf{B}$.

Definition 1.4 — Dot/scalar product. The dot or scalar product of two vectors $\mathbf{A}$ and $\mathbf{B}$ is denoted by $\mathbf{A}\cdot\mathbf{B}$ and is defined to be the real number

$$\mathbf{A}\cdot\mathbf{B}=|\mathbf{A}|\,|\mathbf{B}|\cos\alpha.$$

We can manipulate the expression in the above definition to check its well-posedness: assuming $\mathbf{A},\mathbf{B}$ are non-zero, we find

$$\cos\alpha=\frac{1}{|\mathbf{A}|\,|\mathbf{B}|}\mathbf{A}\cdot\mathbf{B}=\frac{\mathbf{A}}{|\mathbf{A}|}\cdot\frac{\mathbf{B}}{|\mathbf{B}|}=\widehat{\mathbf{A}}\cdot\widehat{\mathbf{B}}.$$

In other words, the angle between two vectors does not depend on the length of the vectors.

R It is legitimate to ask how the expression in the definition above arises: why $\cos\alpha$ and not $\sin\alpha$ etc. The answer is provided later when we discuss coordinates. For now, we note the very useful ensuing properties listed in the following exercise.

Exercise 1.3 Use the definition of the dot product to derive the following properties:
i. $\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A}$;
ii. $\mathbf{A} \cdot \mathbf{A} = |\mathbf{A}|^2$;
iii. $\mathbf{A} \cdot \mathbf{B} = 0 \Leftrightarrow \mathbf{A} \perp \mathbf{B}$.

The scalar product has a straightforward generalisation when we introduce abstract vector spaces. However, this is not the case with the following vector operation: this is included only for completeness.

Cross product

Given two distinct vectors in $\mathbb{E}^3$, their non-parallel direction lines generate uniquely a plane. Various physical models identify phenomena occuring in a direction perpendicular to this plane (e.g., torque, Lorentz force). For this reason, we need to define an operation between two vectors that results in a third vector which is perpendicular to both. This is achieved by the cross product. Before we define this operation, the following convention needs to be introduced.

Definition 1.5 — Right hand rule. An ordered set of **bound vectors** $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ is said to satisfy the **right hand rule** if, given the right hand fingers point in the direction of $\mathbf{A}$ and curve towards $\mathbf{B}$, the thumb points in the direction of $\mathbf{C}$.

R The above definition appears vague: what does it exactly mean that the thumb points in the direction of $\mathbf{C}$? For the purpose of describing the next operation, we will assume that $\mathbf{C}$ is perpendicular to the plane formed by $\mathbf{A}$ and $\mathbf{B}$. Some references specify that the thumb points to the half-space 'above' the plane generated by $\mathbf{A}$ and $\mathbf{B}$ and that $\mathbf{C}$ should be located in this half-space.

Definition 1.6 — Cross-product. The cross-product of two vectors $\mathbf{A}$ and $\mathbf{B}$ is denoted by $\mathbf{A} \times \mathbf{B}$ and is the vector

$$\mathbf{A} \times \mathbf{B} = |\mathbf{A}|\,|\mathbf{B}|\,\hat{\mathbf{n}} \sin\alpha,$$

where $\hat{\mathbf{n}}$ is a unit vector perpendicular to the the plane containing $\mathbf{A}, \mathbf{B}$ with orientation given by the **right-hand rule** (if right-hand fingers point as $\mathbf{A}$ and curve towards $\mathbf{B}$, then thumb points as $\hat{\mathbf{n}}$).

Exercise 1.4 Use the definition of the cross product to derive the following properties:
i. $\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A}$;
ii. $\mathbf{A} \times \mathbf{A} = \mathbf{0}$;
iii. $\mathbf{A} \times \mathbf{B} = \mathbf{0} \Leftrightarrow \mathbf{A} \parallel \mathbf{B}$.

Projections

Projection on a line Given a line d with positive direction given by a unit vector $\widehat{\mathbf{d}}$, we define the projection of a vector $\mathbf{A}$ on the line d as the vector $\mathbf{B}$ given by

$$\mathbf{B} := |\mathbf{A}| \cos\theta \widehat{\mathbf{d}} = (\mathbf{A} \cdot \widehat{\mathbf{d}})\widehat{\mathbf{d}},$$

where θ is the angle between the direction line of $\mathbf{A}$ and d. We also write

$$\mathbf{B} = \mathrm{Proj}_d \mathbf{A} = \mathrm{Proj}_{\widehat{\mathbf{d}}} \mathbf{A}.$$

Projection on a vector Given two vectors $\mathbf{A}, \mathbf{B}$, the projection of $\mathbf{A}$ onto $\mathbf{B}$ is the projection of $\mathbf{A}$ onto the direction line of $\mathbf{B}$, with positive direction given by the direction of $\mathbf{B}$:

$$\mathrm{Proj}_{\mathbf{B}} \mathbf{A} = |\mathbf{A}| \cos\theta \widehat{\mathbf{B}} = \frac{\mathbf{A} \cdot \mathbf{B}}{|\mathbf{B}|^2} \mathbf{B} = \frac{\mathbf{A} \cdot \mathbf{B}}{\mathbf{B} \cdot \mathbf{B}} \mathbf{B},$$

where θ is the angle between the direction lines of $\mathbf{A}$ and $\mathbf{B}$.

1.1.3 Bases

Any vector can be represented as a sum of vectors. A **basis** is defined to be a set of vectors containing the least number of vectors that can be used to represent any given vector. The elements of a basis will be denoted by $\mathbf{e}_i$. In $\mathbb{E}^3$ the least number of vectors needed to represent any vector is 3; thus,

$$\mathbf{A} = A_1\mathbf{e}_1 + A_2\mathbf{e}_2 + A_3\mathbf{e}_3,$$

where $A_i \in \mathbb{R}$ is **the component of $\mathbf{A}$ in the direction $\mathbf{e}_i$**. When the choice of basis is obvious, we will also identify a vector by its components in that basis, written as an ordered triple; thus, we employ the following standard notation:

$$\mathbf{A} = (A_1, A_2, A_3).$$

Orthonormal bases

A convenient choice of basis in $\mathbb{E}^3$ is an orthonormal right-handed basis.

Definition 1.7 — Orthonormal basis. Consider the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, where the basis elements $\mathbf{e}_i$ satisfy for $i, j = 1, 2, 3$,

- $|\mathbf{e}_i| = 1$
- the direction lines of $\mathbf{e}_i$ and $\mathbf{e}_j$ are perpendicular for $i \neq j$.

Then the basis set $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is said to be **orthonormal**.

Since the angles α_{ij} between the direction lines of $\mathbf{e}_i, \mathbf{e}_j$ are either zero ($i = j$) or right angles ($i \neq j$), it follows that

$$\mathbf{e}_i \cdot \mathbf{e}_j = |\mathbf{e}_i|\,|\mathbf{e}_j|\cos\alpha_{ij} = \cos\alpha_{ij} = \delta_{ij} := \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

Definition 1.8 — Right-handed orthonormal set. An orthonormal basis set $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is said to be **right-handed** if

$$\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3.$$

Example 1.1 Cartesian vectors, denoted by $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$, are the most common example of a right-handed orthonormal basis. Using this basis, the position of a point P in 3D is uniquely determined by the distances x, y, z from P to three mutually perpendicular planes. In a Cartesian coordinate system, a point P has an associated position vector bound to O, denoted by $\mathbf{r}_P$ and represented in terms of the Cartesian basis as

$$\mathbf{r}_P = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}.$$

1.1.4 Operations using bases

The use of bases allows for an easy, algebraic evaluation of vector operations. Let $\mathbf{A}, \mathbf{B}$ be vectors in $\mathbb{E}^3$.

Vector addition: this is identified as componentwise addition:

$$\mathbf{A} + \mathbf{B} = (A_1 + B_1)\mathbf{e}_1 + (A_2 + B_2)\mathbf{e}_2 + (A_3 + B_3)\mathbf{e}_3.$$

Equivalently, we write

$$(A_1, A_2, A_3) + (B_1, B_2, B_3) = (A_1 + B_1, A_2 + B_2, A_3 + B_3).$$

Scalar-vector multiplication this is identified as componentwise multiplication:

$$a\mathbf{A} = a(A_1\mathbf{e}_1 + A_2\mathbf{e}_2 + A_3\mathbf{e}_3) = aA_1\mathbf{e}_1 + aA_2\mathbf{e}_2 + aA_3\mathbf{e}_3.$$

Equivalently, we write

$$a(A_1, A_2, A_3) = (aA_1, aA_2, aA_3).$$

Using representations in an orthonormal basis allows for simplified manipulations.

Let $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ be an orthonormal basis for $\mathbb{E}^3$. Then

$$\mathbf{e}_i \cdot \mathbf{e}_j = |\mathbf{e}_i|\,|\mathbf{e}_j| \cos\alpha = 1 \cdot 1 \cdot \cos\alpha = \delta_{ij},$$

since $\cos\alpha$ is non-zero only if $i = j$ (in which case $\alpha = 0$).

Let

$$\mathbf{A} = A_1\mathbf{e}_1 + A_2\mathbf{e}_2 + A_3\mathbf{e}_3, \quad \mathbf{B} = B_1\mathbf{e}_1 + B_2\mathbf{e}_2 + B_3\mathbf{e}_3.$$

Scalar product This simplifies to

$$\mathbf{A} \cdot \mathbf{B} = (A_1\mathbf{e}_1 + A_2\mathbf{e}_2 + A_3\mathbf{e}_3) \cdot (B_1\mathbf{e}_1 + B_2\mathbf{e}_2 + B_3\mathbf{e}_3) = A_1B_1 + A_2B_2 + A_3B_3.$$

Hence,

$$\mathbf{A} \cdot \mathbf{A} = |\mathbf{A}|\,|\mathbf{A}| \cdot 1 = A_1^2 + A_2^2 + A_3^2,$$

so that

$$|\mathbf{A}| = \sqrt{A_1^2 + A_2^2 + A_3^2}.$$

Cross product This simplifies to

$$\mathbf{A} \times \mathbf{B} = (A_2B_3 - A_3B_2)\mathbf{e}_1 + (A_3B_1 - A_1B_3)\mathbf{e}_2 + (A_1B_2 - A_2B_1)\mathbf{e}_3.$$

We write the above expression formally as the following determinant:

$$\mathbf{A} \times \mathbf{B} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{vmatrix}.$$

1.2 Summary

The concepts and definitions outlined in this section arise from a practical need to formalise operations needed in applications. Let us summarise our findings:

- The objects of interest are 3D vectors.
- Vectors are employed in two types of operations: operations involving scalars and vectors, and operations involving vectors only.
- The two operations always return vectors.
- The operations exhibit a range of properties: commutativity, associativity, distributivity etc.
- Some vectors are special in some way (zero vector, unit vector).
- The concept of a basis introduces sets of scalars known as coordinates.
- Using bases, vector operations simplify to elementary scalar computations.

In the following, we aim to define structures known as vector spaces that incorporate and generalise the above observations. One less obvious generalisation concerns the scalars employed. In our discussion, these were just real numbers. We will allow these to be drawn from other sets of numbers, e.g., complex numbers or positive integers chosen in a specific way. The criterion will be that the scalars employed exhibit the same properties as the reals. More precisely, we employ sets of scalars that form a **field** when equipped with certain additive and multiplicative operations.

LECTURE 2

Vector spaces: definitions and examples

2.1 Fields

Before we introduce vector spaces we need to introduce the definition of a field. Fields should be familiar concepts: a typical example is the set of real numbers equipped with the standard operations of addition and multiplication: these allow us to perform calculations and manipulations involving real numbers in a rigourous fashion. Once defined, fields should also feel familiar as they employ the concept (and axioms) of Abelian groups. As you may recall from last year, these are sets endowed with a single operation, which satisfy five properties: closure, associativity, commutativity, the existence of an identity and the existence of an inverse – all with respect to the operation defined on the set. Roughly, speaking, fields are Abelian groups with respect to two operations, with an additional property of distributivity satisfied. The following definition of a field outlines these concepts in a formal way.

Definition 2.1 A **field** is a set $\mathbb{F}$ equipped with binary operations $+$ and $\cdot$, which satisfies the following axioms, for all $a, b, c \in \mathbb{F}$:

FA0 Closure under addition: $a + b \in \mathbb{F}$.
FA1 Associativity of addition: $a + (b + c) = (a + b) + c$.
FA2 Existence of additive identity : For each $a \in \mathbb{F}$, there exists $o \in \mathbb{F}$ such that $a + o = a$.
FA3 Existence of additive inverses: For each $a \in \mathbb{F}$, there exists $a^- \in \mathbb{F}$ such that $a + a^- = o$.
FA4 Commutativity of addition: $a + b = b + a$.
FM0 Closure under multiplication: $a \cdot b \in \mathbb{F}$.
FM1 Associativity of multiplication: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
FM2 Existence of multiplicative identity: There exists $e \in \mathbb{F}$ such that, for all $a \in \mathbb{F}$, $a \cdot e = a$.
FM3 Existence of multiplicative inverses: For each $a \in \mathbb{F}$, $a \neq o$, there exists $a^* \in \mathbb{F}$ such that $a \cdot a^* = e$.
FM4 Commutativity of multiplication: $a \cdot b = b \cdot a$.
FD Distributivity law: $a \cdot (b + c) = a \cdot b + a \cdot c$.

We denote a generic field by $(\mathbb{F}, +, \cdot)$. We will refer to the elements of a field $\mathbb{F}$ as **scalars**; they will be denoted by italic lower-case letters. The corresponding binary operations will be referred to as **scalar addition** and **scalar multiplication**.

R Axioms FA0–FA4 indicate that $(\mathbb{F}, +)$ is an Abelian group when equipped with the addition operation, while axioms FM0–FM4 indicate that $(\mathbb{F} \setminus \{o\}, \cdot)$ is an Abelian group when equipped with the multiplication operation.

R In order to verify that $(\mathbb{F}, +, \cdot)$ is a field, we need to verify the above axioms, which implies establishing the existence of additive and multiplicative identities o and e, as well as that of the additive and multiplicative inverses a^- and a^*.

Example 2.1 The fields $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ and $(\mathbb{C}, +, \cdot)$, where $+, \cdot$ denote the standard addition and multiplication operations have $o = 0, e = 1$. Moreover, given $a \neq 0$, we identify $a^- = -a$ and $a^* = a^{-1}$. Finally, note that $e^- = -1$ and $e^* = 1^{-1} = 1$.

Exercise 2.1 Let $(\mathbb{F}, +, \cdot)$ be a field. Using the notation introduced in the above definition, show that

1. $a^- + a = o$;
2. $e \cdot a = a$;
3. $a^* \cdot a = e$;
4. $(a + b) \cdot c = a \cdot c + b \cdot c$.

While the most common fields employed in this course are $(\mathbb{R}, +, \cdot)$ and $(\mathbb{C}, +, \cdot)$, both of which are infinite, one can also work with finite fields.

Example 2.2 Let p denote a prime number and let $\mathbb{Z}_p$ denote the set of congruence classes modulo p. There are finitely-many elements of $\mathbb{Z}_p$, namely the integers $\{0, 1, \ldots, p-1\}$. Let $\oplus, \odot$ denote addition and multiplication modulo p, i.e.,

$$a \oplus b := (a + b) \bmod p, \quad a \odot b := ab \bmod p.$$

Then $(\mathbb{Z}_p, \oplus, \odot)$ is a (finite) field.

Exercise 2.2 Fill in the following modular arithmetic tables:

$\oplus$	0	1	2	3	4
0					
1					
2					
3					
4					

$\odot$	0	1	2	3	4
0					
1					
2					
3					
4					

Hence, check that $(\mathbb{Z}_5, \oplus, \odot)$ is a field.

2.2 Vector spaces: axioms

We are now in the position to provide a definition for a vector space. This will comprise a set of axioms based on the summary 1.2. In particular, following the example of 3D vectors, we will define two special operations:

- an additive operation involving two vectors (and returning a vector);
- a multiplicative operation involving a scalar and a vector (and returning a vector).

These operations will also need to exhibit certain properties: commutativity, associativity, distributivity. The definition will also require a choice of scalars (i.e., a choice of field). Henceforth, generic vectors will be denoted by bold lower-case letters, unless otherwise indicated.

Definition 2.2 Let $(\mathbb{F}, +, \cdot)$ be a field and let e denote its multiplicative identity. A **vector space** over $\mathbb{F}$ is a set V equipped with the operations of vector addition $\boldsymbol{+}$ and scalar-vector multiplication $\bullet$, satisfying the following axioms for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, $a, b \in \mathbb{F}$:

VA0 Closure under addition: $\mathbf{u} \boldsymbol{+} \mathbf{v} \in V$.
VA1 Associativity of vector addition: $\mathbf{u} \boldsymbol{+} (\mathbf{v} \boldsymbol{+} \mathbf{w}) = (\mathbf{u} \boldsymbol{+} \mathbf{v}) \boldsymbol{+} \mathbf{w}$.
VA2 Existence of vector additive identity: There exists $\mathbf{z} \in V$ such that for all $\mathbf{v} \in V$, $\mathbf{v} \boldsymbol{+} \mathbf{z} = \mathbf{v}$.
VA3 Existence of vector additive inverses: For all $\mathbf{v} \in V$, there exists $\mathbf{v}^- \in V$ such that $\mathbf{v} \boldsymbol{+} \mathbf{v}^- = \mathbf{z}$.
VA4 Commutativity of vector addition: $\mathbf{u} \boldsymbol{+} \mathbf{v} = \mathbf{v} \boldsymbol{+} \mathbf{u}$.
VM0 Closure under scalar-vector multiplication: $a \bullet \mathbf{v} \in V$.
VM1 Associativity of scalar-vector multiplication: $a \bullet (b \bullet \mathbf{v}) = (a \cdot b) \bullet \mathbf{v}$.
VM2 Distributive property of scalar-vector multiplication: $a \bullet (\mathbf{u} \boldsymbol{+} \mathbf{v}) = (a \bullet \mathbf{u}) \boldsymbol{+} (a \bullet \mathbf{v})$.
VM3 Distributive property of scalar addition: $(a + b) \bullet \mathbf{v} = (a \bullet \mathbf{v}) \boldsymbol{+} (b \bullet \mathbf{v})$.
VM4 Multiplicative identity property: $e \bullet \mathbf{v} = \mathbf{v}$, where e is the scalar multiplicative identity in $\mathbb{F}$.

We denote a generic vector space by $(V, \boldsymbol{+}, \bullet, \mathbb{F})$, although often we use the simplified notation $V(\mathbb{F})$, or just V and only specify the field $\mathbb{F}$ (see examples later).

R If axioms VA0–VA4 are satisfied, then $(V, \boldsymbol{+})$ is an Abelian group. The remaining axioms add to this group structure the properties we established for Euclidean vectors (see section 1.2).

R It is important to note that there are **four operations** involved in the above set of axioms: the two field operations $+, \cdot$ and the two operations returning vectors $\boldsymbol{+}, \bullet$. It is evident that these operations only make sense in the given context; however, one tends to be sloppy and use the same symbols for the latter two operations as for the first two. We illustrate this later for the case of a real vector space, which is the most common structure used in this course.

We now turn to some basic properties of vector spaces.

2.3 Vector spaces: basic properties.

Here are some basic properties of vector spaces.

Proposition 2.1 — Elementary properties I. Let $V(\mathbb{F})$ be a vector space. Then the following hold:

1. (Cancellation in sums) For any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, if $\mathbf{u} \boldsymbol{+} \mathbf{v} = \mathbf{u} \boldsymbol{+} \mathbf{w}$ then $\mathbf{v} = \mathbf{w}$.
2. For all $\mathbf{v} \in V$, $o \bullet \mathbf{v} = \mathbf{z}$.
3. For all $\mathbf{v} \in V$, $e^- \bullet \mathbf{v} = \mathbf{v}^-$.
4. For all $a \in \mathbb{F}$, $a \bullet \mathbf{z} = \mathbf{z}$.

Proof. To show 1, we use the axioms VA1–VA4:

$$\mathbf{u} \boldsymbol{+} \mathbf{v} = \mathbf{u} \boldsymbol{+} \mathbf{w} \implies \mathbf{u}^- \boldsymbol{+} (\mathbf{u} \boldsymbol{+} \mathbf{v}) = \mathbf{u}^- \boldsymbol{+} (\mathbf{u} \boldsymbol{+} \mathbf{w}) \implies (\mathbf{u}^- \boldsymbol{+} \mathbf{u}) \boldsymbol{+} \mathbf{v} = (\mathbf{u}^- \boldsymbol{+} \mathbf{u}) \boldsymbol{+} \mathbf{w} \implies \mathbf{z} \boldsymbol{+} \mathbf{v} = \mathbf{z} \boldsymbol{+} \mathbf{w} \implies \mathbf{v} = \mathbf{w}.$$

To show 2, we use axioms VA2, VM4, FA2, VM3 and VM4, in that order:

$$\mathbf{v} \boldsymbol{+} \mathbf{z} = \mathbf{v} = e \bullet \mathbf{v} = (e + o) \bullet \mathbf{v} = e \bullet \mathbf{v} \boldsymbol{+} o \bullet \mathbf{v} = \mathbf{v} \boldsymbol{+} o \bullet \mathbf{v}.$$

Hence, by cancellation in sums, $\mathbf{z} = o \bullet \mathbf{v}$. The third and fourth claims are left as exercises. ■

Note that, by commutativity, if $\mathbf{v} \boldsymbol{+} \mathbf{u} = \mathbf{w} \boldsymbol{+} \mathbf{u}$ then $\mathbf{v} = \mathbf{w}$.

Proposition 2.2 — Elementary properties II. Let $V(\mathbb{F})$ be a vector space. Then:

1. (Cancellation in products)
   (a) Let $o \neq a \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in V$. If $a \bullet \mathbf{u} = a \bullet \mathbf{v}$, then $\mathbf{u} = \mathbf{v}$.
   (b) Let $a, b \in \mathbb{F}$ and $\mathbf{z} \neq \mathbf{u} \in V$. If $a \bullet \mathbf{u} = b \bullet \mathbf{u}$, then $a = b$.

2. Let $a \in \mathbb{F}$ and $\mathbf{u} \in V$. If $a \bullet \mathbf{u} = \mathbf{z}$, then either $a = o$ or $\mathbf{u} = \mathbf{z}$.

Proof. To prove statement 1 we note the following

- Let $a \bullet \mathbf{u} = a \bullet \mathbf{v}$ and $a \neq o$. Since $a \neq o$, by FM3, there exists a^* such that $a^* \cdot a = e$. Multiplying both sides with a^* on the left, we get, using VM1, FM3 and VM4, respectively,

$$a^* \cdot (a \bullet \mathbf{u}) = a^* \cdot (a \bullet \mathbf{v}) \Longrightarrow (a^* \cdot a) \bullet \mathbf{u} = (a^* \cdot a) \bullet \mathbf{v} \Longrightarrow e \bullet \mathbf{u} = e \bullet \mathbf{v} \Longrightarrow \mathbf{u} = \mathbf{v}.$$

- Assume now that $\mathbf{u} \neq \mathbf{z}$. We show the statement by contradiction: assume that $a \neq b$. Then

$$a \bullet \mathbf{u} = b \bullet \mathbf{u} \Longrightarrow a \bullet \mathbf{u} \boldsymbol{+} b^- \bullet \mathbf{u} = b \bullet \mathbf{u} \boldsymbol{+} b^- \bullet \mathbf{u} \Longrightarrow (a + b^-) \bullet \mathbf{u} = o \bullet \mathbf{u} \Longrightarrow (a + b^-) \bullet \mathbf{u} = \mathbf{z}.$$

  Now $a + b^- \neq b + b^- = o$. Hence, there exists $(a + b^-)^*$; multiplying the last identity from the left by this scalar we get $\mathbf{u} = (a + b^-)^* \bullet \mathbf{z} = \mathbf{z}$, which is a contradiction. Hence, we must have $a = b$.

To prove statement 2, we note the following.

- By item 4 in Proposition 2.1 we can write $a \bullet \mathbf{u} = \mathbf{z} = a \bullet \mathbf{z}$; if $a \neq o$, we can use cancellation in products (part (a)) to find $\mathbf{u} = \mathbf{z}$.
- By item 2 in Proposition 2.1 we can write $a \bullet \mathbf{u} = \mathbf{z} = o \bullet \mathbf{u}$; if $\mathbf{u} \neq \mathbf{z}$, we can use cancellation in products (part (b)) to find $a = o$.

Hence, if $a \bullet \mathbf{u} = \mathbf{z}$, then either $a = o$ or $\mathbf{u} = \mathbf{z}$. ■

Exercise 2.3 Show that axioms VA0 and VM0 can be replaced with the requirement that $a \bullet \mathbf{u} \boldsymbol{+} b \bullet \mathbf{v} \in V$ for all $\mathbf{u}, \mathbf{v} \in V$ and for all $a, b \in \mathbb{F}$.

2.4 Examples

We end this lecture with examples of vector spaces, all involving familiar mathematical objects, namely column vectors, polynomials and continuous functions. We also include a generic example of a vector space over a finite fields.

2.4.1 Column vectors

We define (real) column vectors to be vertical arrays of (real) numbers. A generic column vector of size $n \in \mathbb{N}$ has the form

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix},$$

where $v_i \in \mathbb{R}$ $(i = 1, \ldots, n)$. We denote the ith component of $\mathbf{v}$ by $[\mathbf{v}]_i := v_i$; we also write $[v_i]_{1 \le i \le n} = \mathbf{v}$ or simply $[v_i] = \mathbf{v}$. We denote the set of real column vectors of size n by $\mathbb{R}^n$. On this set, we can define the following operations, inspired by the corresponding coordinate operations for 3D vectors:

$$\mathbf{u} \boldsymbol{+} \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \boldsymbol{+} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} := \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix}, \qquad a \bullet \mathbf{v} := \begin{bmatrix} av_1 \\ av_2 \\ \vdots \\ av_n \end{bmatrix} \quad (a \in \mathbb{R}).$$

These operations can be used to generate a vector space structure for $\mathbb{R}^n$, as the following result shows.

Proposition 2.3 Let the set $\mathbb{R}^n$ be equipped with the following operations, for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n, a \in \mathbb{R}$:

i. vector addition: $\mathbf{u} \boldsymbol{+} \mathbf{v} := [u_i + v_i]$;
ii. scalar-vector multiplication: $a \bullet \mathbf{v} := [a \cdot v_i]$.

Then $\mathbb{R}^n$ is a vector space over $\mathbb{R}$.

Proof. The proof is left as an exercise. ■

As part of the above proof, one should identify the zero vector and the additive inverse:

$$\mathbf{z} := \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{v}^- := \begin{bmatrix} -av_1 \\ -av_2 \\ \vdots \\ -av_n \end{bmatrix} =: -\mathbf{v}.$$

The space $\mathbb{R}^n$ is sometimes referred to as Euclidean space, as it can be seen as a generalisation of the sets of Euclidean coordinates in 3D. Note that the convention in this course is to use column vectors as members of $\mathbb{R}^n$, although some references prefer to use row vectors.

2.4.2 Polynomials

Let $\mathcal{P}_n$ denote the set of polynomials of degree at most n with real coefficients:

$$\mathcal{P}_n = \{p(x) = a_0 + a_1 x + \cdots + a_n x^n : a_i \in \mathbb{R}, i = 0, 1, \ldots, n\}.$$

The sum of two polynomials of degree at most n, say p and q, is a polynomial of degree at most n, say r. We write formally

$$r(x) := (p \boldsymbol{+} q)(x) := p(x) + q(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots (a_n + b_n)x^n,$$

given $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ and $q(x) = b_0 + b_1 x + \cdots + b_n x^n$. The scalar multiple of a polynomial of degree at most n is also a polynomial of degree at most n. Formally,

$$(a \bullet p)(x) := a \cdot p(x) = a(a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n) = aa_0 + aa_1 x + \cdots + aa_n x^n.$$

These operations can be used to give $\mathcal{P}_n$ a vector space structure.

Proposition 2.4 Let $\mathcal{P}_n$ be equipped with the following operations, for any $p, q \in \mathcal{P}_n, a \in \mathbb{R}$:

i. Vector addition: $(p \boldsymbol{+} q)(x) := p(x) + q(x)$;
ii. Scalar-vector multiplication: $(a \bullet p)(x) := a \cdot p(x)$.

Then $\mathcal{P}_n$ is a vector space over $\mathbb{R}$.

Proof. The proof is left as an exercise. ■

As part of the above proof, one should identify the 'zero vector' and the additive inverse:

$$\mathbf{z} := 0 + 0x + \cdots + 0x^n, \quad p^-(x) = -a_0 - a_1 x \cdots - a_n x^n.$$

In other words, the zero polynomial is the polynomial with zero coefficients.

2.4.3 Continuous functions

Let $\Omega \subset \mathbb{R}$ and consider the set

$$C(\Omega) = \{f : \Omega \to \mathbb{R} \mid f \text{ is continuous}\}$$

This set is sometimes denoted by $C^0(\Omega)$ (see below for generalisations). We define, as in the case of polynomials,

$$(f \boldsymbol{+} g)(x) := f(x) + g(x), \quad (a \bullet f)(x) = a \cdot f(x).$$

We note that $f \boldsymbol{+} g$ and $a \bullet f$ are also continuous functions.

Proposition 2.5 Let $C(\Omega)$ be equipped with the following operations, for any $f, g \in C(\Omega), a \in \mathbb{R}$:
i. Vector addition: $(f \boldsymbol{+} g)(x) := f(x) + g(x)$;
ii. Scalar-vector multiplication: $(a \bullet f)(x) = a \cdot f(x)$.
Then $C(\Omega)$ is a vector space over $\mathbb{R}$.

Proof. The proof is left as an exercise. ■

As part of the proof, we identify the 'zero vector' as the zero function; the additive inverse of f is $-f$. For smoother functions, one can define the sets $C^k(\Omega)$, where $k \in \mathbb{N}$:

$$C^k(\Omega) = \left\{ f : \Omega \to \mathbb{R} : \frac{\mathrm{d}^j f}{\mathrm{d}x^j} \text{ is continuous}, 0 \le j \le k \right\}.$$

In other words, $C^k(\Omega)$ are the sets of functions that are k times differentiable on Ω for some $k \in \mathbb{N}$. One can similarly check that these are also vector spaces over $\mathbb{R}$, when equipped with the same operations.

2.5 Vector spaces over finite fields

Let $(\mathbb{F}_q, \oplus, \odot)$ denote a finite field with q elements, where q is a prime. Consider the set

$$\mathbb{F}_q^n := \{\mathbf{v} = (v_1, v_2, \ldots, v_n) : v_i \in \mathbb{F}_q\}.$$

Then $\mathbb{F}_q^n$ is a vector space over $\mathbb{F}_q$ when equipped with the vector operations

$$\mathbf{u} \boldsymbol{+} \mathbf{v} := (u_1 \oplus v_1, u_2 \oplus v_2, \ldots u_n \oplus v_n), \quad a \bullet \mathbf{v} = (a \odot v_1, a \odot v_2, \ldots, a \odot v_n),$$

where $a \in \mathbb{F}_q$. Unlike the previous examples, the set $V = \mathbb{F}_q^n$ has a finite number of elements, its cardinality being $\left|\mathbb{F}_q^n\right| = q^n$. A common choice is $\mathbb{F}_q = \mathbb{Z}_q$, with $\oplus, \odot$ denoting addition and multiplication modulo q. These vector spaces are studied in Number Theory and Representation Theory.

2.6 Notation

For the remainder of this course, we will simplify the notation as follows:

- We will use the notation for sets to indicate also the corresponding structures, i.e., fields or vector spaces, e.g., we will write $\mathbb{R}$ to indicate the field $(\mathbb{R}, +, \cdot)$, or $\mathbb{R}^n$ to indicate the vector space $(\mathbb{R}^n, \boldsymbol{+}, \bullet, \mathbb{R})$.
- We will denote all the additive operations by $+$ and all multiplicative operations by $\cdot$ (sometimes even dropping this symbol), e.g., we will write $a \cdot \mathbf{u}$ or $a\mathbf{u}$ instead of $a \bullet \mathbf{u}$.
- We will change notation as follows: $0 = o, 1 = e, -1 = e^-$ and $\mathbf{0} = \mathbf{z}$.

With these simplifications, the definition of a real vector space reads as follows (and as you may find it in other references).

Definition 2.3 A **real vector space** is a set V equipped with two operations, $+$ and $\cdot$, satisfying the following axioms, for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, $a, b \in \mathbb{R}$:
VA0 Closure under vector addition: $\mathbf{u} + \mathbf{v} \in V$.
VA1 Associativity of vector addition: $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.
VA2 Existence of zero vector: There exists $\mathbf{0} \in V$ such that for all $\mathbf{v} \in V$, $\mathbf{v} + \mathbf{0} = \mathbf{v}$.
VA3 Existence of vector additive inverses: For all $\mathbf{v} \in V$, there exists $\mathbf{v}^- \in V$ such that $\mathbf{v} + \mathbf{v}^- = \mathbf{0}$.
VA4 Commutativity of vector addition: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
VM0 Closure under scalar-vector multiplication: $a\mathbf{v} \in V$.
VM1 Associativity of scalar-vector multiplication: $a(b\mathbf{v}) = (ab)\mathbf{v}$.
VM2 Distributive property of scalar-vector multiplication: $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$.
VM3 Distributive property of scalar addition: $(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$.

VM4 Multiplicative identity property: $1\mathbf{v} = \mathbf{v}$.

R The definition of a **complex vector space** is similar, with the scalars a, b drawn from the field $\mathbb{C}$. We will mostly work with real vector spaces, although complex vector spaces will also be considered occasionally.

LECTURE 3

Subspaces

3.1 Characterisation

Any vector space V contains at least one other vector space, assuming they are equipped with the same operations – we will refer to any such space as a subspace of V.

Definition 3.1 Let $(V, +, \cdot, \mathbb{F})$ be a vector space and let U be a subset of V, namely, $U \subseteq V$. If $(U, +, \cdot, \mathbb{F})$ is a vector space, we say that it is a subspace of $(V, +, \cdot, \mathbb{F})$. We write $U(\mathbb{F}) \leq V(\mathbb{F})$, or simply $U \leq V$.

R Note that we use different symbols $\subseteq$ and $\leq$ to establish relations between sets and vector spaces, respectively. Some references use the former symbol to mean both 'subset' and 'subspace'.

The example below considers three special cases of subsets of a given set V.

Example 3.1 Let $V(\mathbb{F})$ be a vector space. We highlight the following three special cases.

- The empty set is a subset of V, but is not a subspace of V since axiom VA2 is not satisfied: the empty set does not contain the zero vector.
- Let $Z = \{\mathbf{0}\}$. Then $Z(\mathbb{F})$ is a vector space, known as **the trivial vector space**. By the above definition, Z is a subspace of any vector space V: $Z \leq V$.
- Since $V \subseteq V$, by the above definition, V is a subspace of V: $V \leq V$.

The example above prompts the following definition.

Definition 3.2 Let $U(\mathbb{F}) \leq V(\mathbb{F})$. If the strict set inclusion $U \subset V$ holds , the subspace $U(\mathbb{F})$ is said to be a **proper subspace** of $V(\mathbb{F})$. We write $U(\mathbb{F}) < V(\mathbb{F})$, or simply $U < V$.

When establishing that a given set V affords a vector space structure, typically we need to check the axioms VA0–VA4 and VM0–VM4. However, in the case of a subset U of V, many of them already hold due to the set inclusion $U \subseteq V$. We can make this precise with the following result.

Proposition 3.1 — Subspace criterion 1: closure criterion. Let $V(\mathbb{F})$ be a vector space. A non-empty subset U of V is a subspace of V over $\mathbb{F}$ if and only if it satisfies the **closure criterion**, i.e., U is closed under vector addition and scalar-vector multiplication.

Proof. Let $V(\mathbb{F})$ be a vector space and let U be a non-empty subset of V.
$\boxed{\Longrightarrow}$ Assume that U is a subspace of V. Then U is a vector space in its own right, i.e., it satisfies the vector space axioms, in particular VA0 and VM0, i.e., U is closed under vector addition and scalar-vector multiplication.
$\boxed{\Longleftarrow}$ Assume that U satisfies the closure criterion. Now, except for VA2 and VA3, all the vector space axioms hold for all the elements of U due to the set inclusion $U \subseteq V$. By closure, $a \bullet \mathbf{u} \in U$ for all $a \in \mathbb{F}$ and all $\mathbf{u} \in U$. Hence, using the elementary properties 2 and 3 in Proposition 2.1,

- VA2 must hold: taking $a = o \in \mathbb{F}$, we find that $\mathbf{z} = o \bullet \mathbf{u} \in U$. Since $\mathbf{u} \in V$, $\mathbf{u} + \mathbf{z} = \mathbf{u}$.
- VA3 must hold: taking $a = e^- \in \mathbb{F}$, we find that $\mathbf{u}^- = e^- \bullet \mathbf{u} \in U$. Since $\mathbf{u}^- \in V$, $\mathbf{u} + \mathbf{u}^- = \mathbf{z}$.

■

Subspace criterion 1 can be recast as follows.

Proposition 3.2 — Subspace criterion 2: linear combination criterion. Let $V(\mathbb{F})$ be a vector space. A non-empty subset U of V is a subspace of V over $\mathbb{F}$ if and only if for any $\mathbf{u}, \mathbf{v} \in U$ and for any $a, b \in \mathbb{F}$, there holds $a\mathbf{u} + b\mathbf{v} \in U$.

Proof. Use the result of Exercise 2.3. ■

R The expression $a\mathbf{u} + b\mathbf{v}$ is called a **linear combination** of the vectors $\mathbf{u}$ and $\mathbf{v}$. Thus, subspace criterion 2 above requires that any linear combination of any two vectors in U should also be in U.

3.2 Examples

Let us consider subspaces of the three vector space examples presented in Lecture 2.

Example 3.2 — Column vectors. Let $V = \mathbb{R}^3$ and define the subset $U \subset V$

$$U = \left\{ \begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix} : v_1, v_2 \in \mathbb{R} \right\}.$$

Given any two vectors $\mathbf{u}, \mathbf{v} \in U$, there holds for any $a, b \in \mathbb{R}$,

$$a\mathbf{u} + b\mathbf{v} = a \begin{bmatrix} u_1 \\ u_2 \\ 0 \end{bmatrix} + b \begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix} = \begin{bmatrix} au_1 + bv_1 \\ au_2 + bv_2 \\ 0 \end{bmatrix} \in U,$$

so that, by Subspace criterion 2 (Proposition 3.2), U is a subspace of V.

Remark This example can be extended to $V = \mathbb{R}^n$, with $U \subset V$ being the set of vectors with $k < n$ non-zero entries in fixed locations.

Example 3.3 — Polynomials. Recall that $\mathcal{P}_n$ is the space of polynomials of degree at most n:

$$\mathcal{P}_n = \{p(x) = a_0 + a_1 x + \cdots + a_n x^n : a_i \in \mathbb{R}, i = 0, 1, \ldots, n\}.$$

Let $k < n$. Then $\mathcal{P}_k \subset \mathcal{P}_n$ is a subspace of $\mathcal{P}_n$ when equipped with the operations of polynomial addition and multiplication of polynomials by scalars.

The above two examples are related: since any polynomial $p_n(x)$ of degree n can be uniquely identified by the vector of its coefficients:

$$p_n(x) \longleftrightarrow [a_0, a_1, \cdots, a_n],$$

we find, given $p_k \in \mathcal{P}_n$,

$$p_k(x) \longleftrightarrow [a_0, a_1, \cdots, a_k, 0, \cdots, 0].$$

The latter row vector contains $k < n$ non-zero entries: see final remark in Example 3.2

Example 3.4 — Continuous functions. Let $V = C^0(\Omega)$ denote the space of functions $f : \Omega \to \mathbb{R}$ which are continuous. The space $U = C^1(\Omega)$ was defined as the space of continuous functions which are differentiable $k = 1$ times. Since a differentiable function is continuous, $U = C^1(\Omega) \subset C^0(\Omega) = V$. One can immediately use Proposition 3.2 to show that U is a subspace of V.

We consider now two set operations which preserve the vector space structure: intersection and sum.

3.3 Intersection of subspaces

The intersection of subspaces is also a subspace, as the following result shows.

Theorem 3.3 — Intersection of subspaces. Let $U_1(\mathbb{F}), U_2(\mathbb{F})$ be subspaces of $V(\mathbb{F})$. Then the set intersection $U_1 \cap U_2$ is also a subspace of V when equipped with the vector operations defined on V.

Proof. Let $U_1, U_2 \leq V$. Since $\mathbf{0} \in U_1$ and $\mathbf{0} \in U_2$, the set intersection $U_1 \cap U_2$ is always non-empty. Let $\mathbf{u}, \mathbf{v} \in U_1 \cap U_2$. Then $\mathbf{u}, \mathbf{v} \in U_1$ and also $\mathbf{u}, \mathbf{v} \in U_2$. Since U_1, U_2 are subspaces, they satisfy the closure criterion, so that $a\mathbf{u} + b\mathbf{v} \in U_1$ and also $a\mathbf{u} + b\mathbf{v} \in U_2$. Hence $a\mathbf{u} + b\mathbf{v} \in U_1 \cap U_2$ and by Proposition 3.2 $U_1 \cap U_2$ is a subspace of V. ■

R It is a standard convention to denote the intersection of subspaces using the same symbol as that for intersection of sets, with the vector space operations and the field assumed to be clear from the context. For example, the result of the above theorem can be written as $U_1(\mathbb{F}) \cap U_2(\mathbb{F}) \leq V(\mathbb{F})$, or simply $U_1 \cap U_2 \leq V$.

Note that the above result can be immediately generalised to the intersection of several subsets of V:

$$\bigcap_{i=1}^{m} U_i \leq V \qquad (U_i \leq V).$$

Example 3.5 Consider the subspaces $U_1, U_2 \leq \mathbb{R}^2$:

$$U_1 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x = y \right\}, \quad U_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x = -y \right\}.$$

Then $U_1 \cap U_2 = \{\mathbf{0}\}$. Note that both sets can be identified as lines in the plane through the origin.

3.4 Sum of subspaces

Recall that the sum of two sets U_1, U_2 is the set W given by

$$W := U_1 + U_2 := \{\mathbf{u}_1 + \mathbf{u}_2 : \mathbf{u}_1 \in U_1, \mathbf{u}_2 \in U_2\}.$$

Theorem 3.4 — Sum of subspaces. Let $U_1(\mathbb{F}), U_2(\mathbb{F})$ be subspaces of $V(\mathbb{F})$. Then the set sum $U_1 + U_2$ is also a subspace of V when equipped with the vector operations defined on V.

Proof. The proof is left as an exercise. ■

As before, the above result can be immediately generalised to the sum of several subsets of V:

$$\sum_{i=1}^{m} U_i \leq V \qquad (U_i \leq V).$$

Example 3.6 Let $V = \mathbb{R}^3$ and consider the following subsets of V:

$$U_1 = \left\{ \begin{bmatrix} u_1 \\ 0 \\ 0 \end{bmatrix} : u_1 \in \mathbb{R} \right\}, \qquad U_2 = \left\{ \begin{bmatrix} 0 \\ u_2 \\ 0 \end{bmatrix} : u_2 \in \mathbb{R} \right\}.$$

Then $W = U_1 + U_2$ is the set

$$W = \left\{ \begin{bmatrix} u_1 \\ u_2 \\ 0 \end{bmatrix} : u_1, u_2 \in \mathbb{R} \right\},$$

which is indeed a subspace of V (see Example 3.2).

In the above example, every $\mathbf{w} \in W$ can be written **uniquely** as a sum of elements in U_1 and U_2. Here is an example where this is not the case.

Example 3.7 Let $V = \mathbb{R}^3$ and consider the following subsets of V:

$$U_1 = \left\{ \begin{bmatrix} u_1 \\ u_2 \\ 0 \end{bmatrix} : u_1, u_2 \in \mathbb{R} \right\}, \qquad U_2 = \left\{ \begin{bmatrix} u_3 \\ 0 \\ 0 \end{bmatrix} : u_3 \in \mathbb{R} \right\}.$$

We note that

$$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

where the first element in each sum is in U_1, with the second in U_2.

These examples indicate that the sum of subspaces should be investigated further: this gives rise to the concept of direct sum.

3.4.1 Direct sum of subspaces

A special case of a sum of subspaces is that of a direct sum. This concept is described in the following definition.

Definition 3.3 — Direct sum. Let $U_1, U_2 \leq V$. Then the subspace $W = U_1 + U_2$ is said to be the direct sum of U_1 and U_2 if for every non-zero $\mathbf{w} \in W$ there exist unique $\mathbf{u}_1 \in U_1$ and $\mathbf{u}_2 \in U_2$ such that $\mathbf{w} = \mathbf{u}_1 + \mathbf{u}_2$. In this case, we write $W = U_1 \oplus U_2$.

The main interest in introducing the concept of direct sum concerns the representation of a vector space as a direct sum of subspaces U_i

$$V = U_1 \oplus U_2 \oplus \cdots \oplus U_n = \bigoplus_{i=1}^{n} U_i.$$

In this case, the subspaces can be viewed as 'building blocks' for V. The uniqueness mentioned in the definition relates to the concept of basis which we will encounter later. For now, let us establish ways of identifying direct sums.

Proposition 3.5 — Direct sum criterion 1: trivial intersection. Let $U_1(\mathbb{F}), U_2(\mathbb{F}) \leq V(\mathbb{F})$. Then the subspace $W = U_1 + U_2$ is a direct sum if and only if $U_1 \cap U_2 = \{\mathbf{0}\}$.

Proof. Let $U_1(\mathbb{F}), U_2(\mathbb{F}) \leq V(\mathbb{F})$ and define $W = U_1 + U_2$.
$\boxed{\implies}$ Let W be a direct sum: $W = U_1 \oplus U_2$. Assume, by contradiction, that there exists a non-zero $\mathbf{w}$ such that $\mathbf{w} \in U_1 \cap U_2$. Since the intersection is a vector space, we must have $-\mathbf{w} \in U_1 \cap U_2$. Choosing first $\mathbf{w} \in U_1, -\mathbf{w} \in U_2$ and then $-\mathbf{w} \in U_1, \mathbf{w} \in U_2$, we can write

$$\mathbf{0} = \mathbf{w} + (-\mathbf{w}) = (-\mathbf{w}) + \mathbf{w},$$

which contradicts the fact that W is a direct sum (due to the non-unique representation of $\mathbf{0}$ as sum of elements in U_1, U_2). The only possibility is that $\mathbf{w} = \mathbf{0}$, which is a contradiction.
$\boxed{\impliedby}$ Let now $U_1 \cap U_2 = \{\mathbf{0}\}$. We show that $W = U_1 \oplus U_2$ by contradiction. Assume that there exists $\mathbf{w} \in W$ which can be written non-uniquely in two distinct ways as

$$\mathbf{w} = \mathbf{u}_1 + \mathbf{u}_2 = \mathbf{u}_1' + \mathbf{u}_2',$$

where $\mathbf{u}_1, \mathbf{u}_1' \in U_1, \mathbf{u}_2, \mathbf{u}_2' \in U_2$, with $\mathbf{u}_1 \neq \mathbf{u}_1'$ and/or $\mathbf{u}_2 \neq \mathbf{u}_2'$. We can now manipulate the above relation:

$$\mathbf{u}_1 - \mathbf{u}_1' = \mathbf{u}_2' - \mathbf{u}_2 =: \mathbf{v}$$

and since $\mathbf{u}_1 - \mathbf{u}_1' \in U_1$ and $\mathbf{u}_2' - \mathbf{u}_2 \in U_2$ we must have $\mathbf{v} \in U_1 \cap U_2 = \{\mathbf{0}\}$. Hence $\mathbf{v} = \mathbf{0}$ and therefore $\mathbf{u}_1 = \mathbf{u}_1'$ *and* $\mathbf{u}_2 = \mathbf{u}_2'$, which is a contradiction. ■

R The previous result confirms that direct sums allow us to represent a vector space as sums of almost disjoint sets U_1 and U_2, with the only common element being zero. This type of representation of vectors in a vector space is very useful in practice as it avoids 'redundancy'; it also arises quite naturally in contexts that we will consider in the second half of this course.

We end with the following result which describes an alternative characterisation of direct sums.

Proposition 3.6 — Direct sum criterion 2: trivial zero sum. Let $U_1(\mathbb{F}), U_2(\mathbb{F}) \leq V(\mathbb{F})$. Then $W = U_1 + U_2$ is a direct sum if and only if the zero representation given by $\mathbf{0} = \mathbf{u}_1 + \mathbf{u}_2$ implies $\mathbf{u}_1 = \mathbf{u}_2 = \mathbf{0}$.

Proof. Let $U_1(\mathbb{F}), U_2(\mathbb{F}) \leq V(\mathbb{F})$ and define $W = U_1 + U_2$.
$\boxed{\implies}$ Let W be a direct sum: $W = U_1 \oplus U_2$. Assume by contradiction that $\mathbf{0} = \mathbf{u}_1 + \mathbf{u}_2$, with $\mathbf{u}_1 \neq \mathbf{0} \neq \mathbf{u}_2$. By definition of direct sum, this should be the only representation of $\mathbf{0}$. However, $\mathbf{0} = \mathbf{0} + \mathbf{0}$, with $\mathbf{0} \in U_1$ and $\mathbf{0} \in U_2$. This is a contradiction.
$\boxed{\impliedby}$ Let the zero representation $\mathbf{0} = \mathbf{u}_1 + \mathbf{u}_2$ imply $\mathbf{u}_1 = \mathbf{u}_2 = \mathbf{0}$. Assume by contradiction that W is not a direct sum. Then there exists $\mathbf{w} \in W$ which can be written non-uniquely in two distinct ways as

$$\mathbf{w} = \mathbf{u}_1 + \mathbf{u}_2 = \mathbf{u}_1' + \mathbf{u}_2',$$

where $\mathbf{u}_1, \mathbf{u}_1' \in U_1, \mathbf{u}_2, \mathbf{u}_2' \in U_2$, with $\mathbf{u}_1 \neq \mathbf{u}_1'$ and/or $\mathbf{u}_2 \neq \mathbf{u}_2'$. Then, taking the difference we get

$$\mathbf{0} = (\mathbf{u}_1 - \mathbf{u}_1') + (\mathbf{u}_2 - \mathbf{u}_2') = \mathbf{u}_1'' + \mathbf{u}_2'',$$

where, by closure, $\mathbf{u}_1'' \in U_1$ and $\mathbf{u}_2'' \in U_2$. Since this is a representation of zero, by our assumption we must have $\mathbf{u}_1'' = \mathbf{u}_2'' = \mathbf{0}$, which in turn implies that $\mathbf{u}_1 = \mathbf{u}_1'$ *and* $\mathbf{u}_2 = \mathbf{u}_2'$, which is a contradiction. ■

LECTURE 4

Spanning sets

4.1 Linear combinations. Span

Subspaces can be defined using the concept of linear combinations. We have already come across these when discussing subspace criterion 2 (see remark after Proposition 3.2). Given their obvious relevance, linear combinations can be defined more generally.

Definition 4.1 — Linear combination. Let $V(\mathbb{F})$ be a vector space and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be vectors in V. A linear combination of these vectors is any vector $\mathbf{v}$ of the form

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k,$$

where $a_1, a_2, \ldots, a_k \in \mathbb{F}$.

Any vector in V can be expressed **non-uniquely** as a linear combination of some vectors in V; for example, in the above definition one can choose $a_1 = 1, \mathbf{v}_1 = \mathbf{v}$ and $a_i = 0, i = 2, \ldots, k$. We also note here the special case of the zero vector which can be seen to be **the trivial linear combination** with all scalars a_i equal to zero.

Let us consider now the case of a linear combination that involves a single vector. Let V be a vector space over $\mathbb{F}$ and let $\mathbf{v} \in V$ be given. Then $a\mathbf{v} \in V$ for any $a \in \mathbb{F}$, so that we can identify the following subset of V:

$$U := \{a\mathbf{v} : a \in \mathbb{F}\}.$$

The set U contains just multiples of the given vector $\mathbf{v}$; in this sense, it can be seen as being entirely determined, or generated, by $\mathbf{v}$. We say that U is spanned by $\mathbf{v}$ and write

$$U = \operatorname{span}\{\mathbf{v}\} := \{a\mathbf{v} : a \in \mathbb{F}\}.$$

One can show that U is a subspace, as the exercise below indicates.

Exercise 4.1 Let V be a vector space and let $\mathbf{v} \in V$. Show that $\operatorname{span}\{\mathbf{v}\}$ is a subspace of V.

Can we use more than one vector to define a span? Are spans always subspaces? The answer to the both questions is yes: first, let us generalise the concept of span with the following definition.

Definition 4.2 — Span of a finite set. Let $V(\mathbb{F})$ be a vector space and let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ denote a non-empty finite set of vectors in V. The span of S is the set of all linear combinations of vectors in S:

$$\text{span}S := \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} := \{a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k : a_i \in \mathbb{F}, i = 1, 2, \ldots, k\}.$$

If $S = \emptyset$, we define $\text{span}S = \{\mathbf{0}\}$.

R One may argue that the concept of span when S is empty does not make sense. A possible justification is given by the following standard 'computer' evaluation of a sum using a *for loop*:

$sum = \mathbf{0}$ (set the sum to be zero initially)
for $i = 1 : |S|$
$\quad sum = sum + a_i\mathbf{v}_i$;
end

In the case $S = \emptyset$, the loop will not run, as $|S| = 0$, so the calculation will return the zero vector.

Let us consider some examples to get a feel for the concept of span.

Example 4.1 — Span of column vectors (1). Consider the following vectors in $\mathbb{R}^3$:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Then

$$\text{span}\{\mathbf{e}_1, \mathbf{e}_2\} = \{a_1\mathbf{e}_1 + a_2\mathbf{e}_2 : a_1, a_2 \in \mathbb{R}\} = \left\{ \begin{bmatrix} a_1 \\ a_2 \\ 0 \end{bmatrix} : a_1, a_2 \in \mathbb{R} \right\},$$

which is a subspace of $\mathbb{R}^3$ (see Example 3.2).

Example 4.2 — Span of column vectors (2). Consider the following vectors in $\mathbb{R}^3$:

$$\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}.$$

Then

$$\text{span}\{\mathbf{u}, \mathbf{v}\} = \{a\mathbf{u} + b\mathbf{v} : a, b \in \mathbb{R}\} = \left\{ \begin{bmatrix} a+b \\ b \\ 0 \end{bmatrix} : a, b \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} a' \\ b \\ 0 \end{bmatrix} : a', b \in \mathbb{R} \right\},$$

while

$$\text{span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\} = \{a\mathbf{u} + b\mathbf{v} + c\mathbf{w} : a, b, c \in \mathbb{R}\} = \left\{ \begin{bmatrix} a+b+c \\ b+2c \\ 0 \end{bmatrix} : a, b, c \in \mathbb{R} \right\} =: \left\{ \begin{bmatrix} a' \\ b' \\ 0 \end{bmatrix} : a', b' \in \mathbb{R} \right\}.$$

Both spans therefore represent the same subspace of $\mathbb{R}^3$ as in the previous example.

At this stage, it is natural to ask some 'obvious' questions:

- Is a span a subspace?
- Is a subspace a span?
- Can we 'build up' a vector space V using spans?

- Can we characterise the intersection/sum of subspaces in terms of some/any spans?

The following results answers the first two questions.

Proposition 4.1 — A span is a subspace. Let $V(\mathbb{F})$ be a vector space and let $S \subseteq V$. Then span$S$ is a subspace of $V(\mathbb{F})$.

Proof. If $S = \emptyset$, then, by definition, span$S = \{\mathbf{0}\}$. By Exercise 4.1, this is a subspace of $V(\mathbb{F})$. Assume now that S is a non-empty set of vectors in V: $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$. Let $\mathbf{u}, \mathbf{v} \in \text{span}S$. By definition,

$$\mathbf{u} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k, \qquad \mathbf{v} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_k\mathbf{v}_k,$$

for some $a_i, b_i \in \mathbb{F}$, $i = 1, 2, \dots k$. Then

$$\begin{aligned} a\mathbf{u} + b\mathbf{v} &= a(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k) + b(b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_k\mathbf{v}_k) \\ &=: c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k \in \text{span}S, \end{aligned}$$

where $c_i := aa_i + bb_i \in \mathbb{F}$ for $i = 1, 2, \dots k$. By the subspace criterion 2 (Proposition 3.2), span$S$ is a subspace of V. ∎

Before we answer the second question above, we need to make precise the notion of a **span of an infinite set**. More precisely, if the set S in Definition 4.2 is infinite, then we might be tempted to consider linear combinations which involve all the elements in S. However, not all infinite sums of vectors are well defined, as the following example shows.

Example 4.3 Let $V = \mathbb{R}^2$ and consider the following infinite set of vectors in V:

$$S = \left\{ \mathbf{v}_k := \begin{bmatrix} k \\ 0 \end{bmatrix} : k \in \mathbb{N} \right\}.$$

Then

$$\sum_{k=1}^{\infty} \mathbf{v}_k = \begin{bmatrix} \sum_{k=1}^{\infty} k \\ 0 \end{bmatrix} \notin V,$$

since the first entry is not a real number. Note, however, that any finite sum yields a vector in V.

The above example indicates how we can generalise our definition of span to the case where the set is infinite.

Definition 4.3 — Span of an infinite set. Let $V(\mathbb{F})$ be a vector space and let S denote an infinite set of vectors in V. The span of S is the set of all finite linear combinations of vectors in S.

R It is evident that the above definition coincides with Definition 4.2 when S is a finite set. However, it is important to highlight this distinction by giving two separate definitions.

Proposition 4.2 — A subspace is a span. Let $U(\mathbb{F})$ be a subspace of $V(\mathbb{F})$. Then $U = \text{span}U$.

Proof. To show this set equality, we need to show that $U \subseteq \text{span}U$ and $\text{span}U \subseteq U$. By closure, any finite linear combination of elements of U is in U. Hence $\text{span}U \subseteq U$. However, there also holds (trivially, by the definition of span) that $U \subseteq \text{span}U$. Hence, $U = \text{span}U$. ∎

Note that since any vector space is its own subspace, we have that $V = \text{span}V$. The remaining questions above can also be answered in the affirmative; the details will be provided in the next lecture. We end this discussion with one last question: how do we establish if a non-zero vector is in a given span? The answer is: via a calculation. We illustrate this with an example.

Example 4.4 Let $U = \text{span}\{\mathbf{u}, \mathbf{w}\}$, where

$$\mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} 0 \\ 4 \\ 3 \end{bmatrix}.$$

A non-zero vector $\mathbf{v}$ is in U if it can be written as a linear combination of $\mathbf{u}$ and $\mathbf{w}$:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in U \iff \mathbf{v} = a\mathbf{u} + b\mathbf{w} \iff \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = a \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + b \begin{bmatrix} 0 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} a \\ a+4b \\ a+3b \end{bmatrix},$$

where a, b are not both zero. This holds provided that $a = v_1$ and $b = v_2 - v_3$ and that $v_1 \neq 0, v_2 \neq v_3$.

4.2 Spanning sets

We have seen how a set of vectors can be used to generate other vectors as linear combinations. Given $S \subset V$, we have that $\text{span} S \subseteq V$. A natural question to ask is when do we have $\text{span} S = V$? In Example 5.5, one can establish that

$$\text{span}\{(1,2),(2,1)\} = \{(0,0),(1,2),(2,1)\},$$

while (check this)

$$\text{span}\{(1,2),(1,1)\} = \mathbb{Z}_3^2.$$

In other words, one can generate the whole set $\mathbb{Z}_3^2$ using two elements, provided they are chosen suitably. This is an example of a generating set (otherwise known as a **spanning set**) for $\mathbb{Z}_3^2$.

Definition 4.4 — Spanning set. Let $V(\mathbb{F})$ be a vector space and let S be a set of vectors in V. We say S is a spanning set for V if $\text{span} S = V$.

R Since $S \subseteq V$, if follows by closure that $\text{span} S \subseteq V$. Hence, to check the set equality $\text{span} S = V$ we only need to check that $V \subseteq \text{span} S$.

Note that every non-trivial vector space has at least one spanning set since $\text{span} V = V$. While they always exist, spanning sets are not unique, with some having considerably fewer vectors than the set V itself.

Example 4.5 The set $V = \mathbb{R}^3$ contains infinitely-many vectors. The following sets contain finitely-many vectors and are both spanning sets for $\mathbb{R}^3$:

$$S_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}, \quad S_2 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

4.3 Minimal spanning sets

Note that if we remove any vector from S_1 in the above example, then the set is not a spanning set for V anymore, while if we remove any vector from S_2, then the resulting set is still a spanning set for V. These observations seem to suggest that one may be able to trim down a spanning set, but up to a point! To make this more precise, let us provide the following definition. Note that the statement considers separately the case where $V(\mathbb{F})$ is the trivial vector space, i.e., $V = \{\mathbf{0}\}$ and a generic non-trivial vector space.

Definition 4.5 — Minimal spanning set. Let S be a spanning set for a vector space $V(\mathbb{F})$.
i. If $V(\mathbb{F})$ is trivial, the minimal spanning set if defined to be the empty set.
ii. If $V(\mathbb{F})$ is non-trivial, S is called **a minimal spanning set** if

$$\text{span} S \setminus \{\mathbf{v}\} \subset V \quad \text{for any } \mathbf{v} \in S.$$

In other words, S is a minimal spanning set if removing any of its elements also removes the spanning property of S. This is indeed possible as the above example shows in the case of S_1.

R The above definition considers separately the case where $V(\mathbb{F})$ is the trivial vector space, as we want to exclude the possibility that S is the empty set (see Definition 4.2); in this case there is no $\mathbf{v}$ in S and the statement in the above definition does not make sense.

At this stage, the usual existence and uniqueness questions arise:
- does a minimal spanning set always exist?
- if a minimal spanning set exists, is it unique?

We answer the first question for the case where the spanning set is a finite set. The answer to the second question is in the negative (see counter-example below).

Proposition 4.3 Every vector space $V(\mathbb{F})$ has a minimal spanning set.

Proof. If $V(\mathbb{F})$ is trivial, the minimal spanning set exists: it is the empty set, by the above definition. Assume now that $V(\mathbb{F})$ is non-trivial. Let S be a spanning set for $V(\mathbb{F})$ and let $k := |S| \in \mathbb{N}$. Define $S_k := S$. We consider two cases.
i. $k = 1$. In this case, $S_1 = \{\mathbf{v}\}$ for some vector $\mathbf{v} \in V$ and this is a minimal spanning set, since $S_0 := S_1 \setminus \{\mathbf{v}\} = \emptyset$ and $\text{span} S_0 = \{\mathbf{0}\} \subset \text{span} S_1 = V$.
ii. $k > 1$. Let $\mathbf{v} \in S_k$ and define $S_{k-1} := S_k \setminus \{\mathbf{v}\}$. Then $\text{span} S_{k-1} \subseteq \text{span} S_k = V$. If $\text{span} S_{k-1} \subset V$, then, by definition, S_k is a minimal spanning set. Otherwise, $\text{span} S_{k-1} = V$ and S_{k-1} is a spanning set for $V(\mathbb{F})$, with $|S_{k-1}| = k - 1$. We can now repeat the argument for S_{k-1}, S_{k-2} and so on, to deduce that S_r is a minimal spanning set for some $r \leq k$ (and with $r > 1$).

■

The previous result is related to the following property of spanning sets.

Corollary 4.4 Every spanning set contains a minimal spanning set.

To see that minimal spanning sets are not unique, consider the following examples.

Example 4.6 Let $S_1 = \{1, x\}$ and $S_2 = \{1, 1 + x\}$. Then $\text{span} S_1 = \text{span} S_2 = \mathcal{P}_1$ and both S_1 and S_2 are minimal spanning sets for $\mathcal{P}_1$, as we cannot remove any elements from either without changing the span.

Example 4.7 Let

$$S_1 = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}, \quad S_2 = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \end{bmatrix} \right\}.$$

It is straightforward to see that both S_1 and S_2 are distinct minimal spanning sets for $V = \mathbb{R}^2$.

The concept of a spanning set S is useful in that it allows us to represent any vector in V as a linear combination of vectors in S, where the span employs the vector space operations associated with $V(\mathbb{F})$. Minimal spanning sets provide this representation in the most 'economical' way. Given a certain choice of minimial spanning set, is this representation unique? We provide an answer to this question in the next lecture.

LECTURE 5

Linearly independent sets

5.1 Finite-dimensional vector spaces

We discuss the concept of dimension later. However, at this stage we can provide some pointers based on the discussion regarding infinite spanning sets from the previous lecture.

Definition 5.1 A vector space is said to be **finite-dimensional** if it has a finite spanning set. If there is no finite spanning set, the vector space is said to be **infinite-dimensional**.

Here is a typical example of an infinite-dimensional space.

Example 5.1 Let $V = \mathcal{P}(\mathbb{R})$ denote the vector space of polynomials of arbitrary degree with real coefficients. This vector space is infinite-dimensional. To see this, assume that there exists a finite spanning set containing polynomials $p_i(x)$, $1 \leq i \leq k$:

$$S = \{p_1(x), p_2(x), \ldots, p_k(x)\}.$$

Then

$$\operatorname{span} S = \{p(x) := a_1 p_1(x) + a_2 p_2(x) + \cdots + a_k p_k(x) : a_i \in \mathbb{R}\},$$

where

$$\deg p = \max_{1 \leq i \leq k} \deg p_i =: n.$$

Then $x^{n+1} \in \mathcal{P}(\mathbb{R})$, but $x^{n+1} \notin \operatorname{span} S$, which is a contradiction. In other words, for any choice of polynomials p_i and for any finite k (i.e., for any finite set S of vectors in V), there will always exist a polynomial which is not in the span of S. Hence, V must be infinite-dimensional.

Infinite-dimensional vector spaces are important in many areas of mathematics; they will certainly arise in courses you may study later. However, the focus of this course, and of Linear Algebra generally, is on finite-dimensional vector spaces. As such, the remainder of these notes will assume that spanning sets are finite sets and vector spaces are finite-dimensional.

5.2 Linear independence

The three examples of spans of column vectors in the previous lecture indicate that

- different sets of vectors can yield the same span;
- the same subspace can be spanned by sets with different cardinality.

The latter observation suggests that there may exist in general a representation of a subspace that employs the least number of vectors. This is indeed the case: sets with 'unnecessary' vectors in this context are known as linearly dependent sets. To make this more precise, consider the following result.

Proposition 5.1 Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ and let $\mathbf{u} \in \text{span}S$. Then

$$\text{span}S = \text{span}S \cup \{\mathbf{u}\}\,.$$

If $\mathbf{u} \in \text{span}S$, there holds

$$\mathbf{u} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k \iff a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k + (-1)\mathbf{u} = \mathbf{0}.$$

In other words, $\mathbf{u}$ can be expressed in terms of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$, and is thus dependent on these vectors, provided the expression on the right holds. This observation justifies the following definition.

Definition 5.2 — Linearly dependent set. Let $V(\mathbb{F})$ be a vector space and let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ denote a set of vectors in V. We say S is linearly dependent provided there exist scalars $a_1, a_2, \dots, a_k \in \mathbb{F}$ not all zero such that

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k = \mathbf{0}. \tag{5.1}$$

R Another interpretation of (5.1) is that the zero vector can be written as a non-trivial linear combination of the vectors in S.

Definition 5.3 — Linearly independent set. A set of vectors that is not linearly dependent is called linearly independent.

A less intuitive example employing this definition is contained in the following result.

Proposition 5.2 The empty set is a linearly independent set.

Proof. If $S = \emptyset$, there are no vectors that can be employed to construct a linear combination equal to the zero vector. Hence the empty set is not linearly dependent. ■

When a set is non-empty, one can rephrase the definition of linear independence as the logical complement of Definition 5.2 (see also the remark following it).

Definition 5.4 — Linearly independent set. Let $V(\mathbb{F})$ be a vector space and let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ denote a non-empty set of vectors in V. We say S is linearly independent provided the zero vector can only be written as the trivial linear combination of vectors in S. In other words,

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k = \mathbf{0} \implies a_1 = a_2 = \cdots = a_k = 0. \tag{5.2}$$

The following results concern some special cases.

Proposition 5.3 Any set S containing the zero vector is linearly dependent.

Proof. Let $\mathbf{v} := \mathbf{0} \in S$. Then $1\mathbf{v} = \mathbf{0}$, so that $\mathbf{0}$ can be written as a non-trivial linear combination of vectors in S. ■

Corollary 5.4 Let $V(\mathbb{F})$ be a vector space. Then V is a linearly dependent set.

Proof. Since $V(\mathbb{F})$ is a subspace, there holds $\mathbf{0} \in V$. By Proposition 5.3, the set V is linearly dependent. ■

Proposition 5.5 A set S containing a single non-zero vector is linearly independent.

Proof. Let $\mathbf{0} \neq \mathbf{v} \in S$. By Proposition 2.2, $a\mathbf{v} = \mathbf{0} \implies a = 0$, so that S is linearly independent. ■

Let us examine some examples of linear dependence and independence.

Example 5.2 Consider again the following vectors in $\mathbb{R}^3$:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

We note that

$$\mathbf{v}_3 = \frac{1}{2}\mathbf{v}_1 + \frac{1}{2}\mathbf{v}_2,$$

so that the set $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent. On the other hand, and two vectors of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are linearly independent. For example,

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 = \mathbf{0} \implies \begin{bmatrix} a_1 \\ 2a_1 \\ 0 \end{bmatrix} + \begin{bmatrix} a_2 \\ 0 \\ 0 \end{bmatrix} \implies \left\{ \begin{aligned} a_1 + a_2 &= 0 \\ 2a_1 &= 0 \end{aligned} \right. \implies a_1 = a_2 = 0.$$

Example 5.3 Let $S = \{a_0 + a_1 x : a_0, a_1 \in \mathbb{R}\}$. Then $S(\mathbb{R})$ is a vector space, as it is the set of polynomials of degree at most one: $S(\mathbb{R}) = \mathcal{P}_1(\mathbb{R})$. By Corollary 5.4, the set S is linearly dependent. On the other hand, the set $S = \{1, x\}$ is linearly independent, since

$$a_1 1 + a_2 x = \mathbf{0} \implies a_1 + a_2 x = \mathbf{0} \implies p = \mathbf{0} \implies a_1 = a_2 = 0,$$

where we defined p to be the polynomial $p(x) = a_1 + a_2 x$. Note that while we tried to use the formalism from Lecture 2, in general, one could also view the statement on the left as an identity that has to be satisfied for any real value assigned to x, e.g., for $x = 1$ and $x = 2$ we obtain a linear system in a_1, a_2:

$$a_1 + a_2 x = 0 \implies \left[\begin{aligned} a_1 + a_2 &= 0 \\ a_1 + 2a_2 &= 0 \end{aligned} \right. \implies a_1 = a_2 = 0.$$

Example 5.4 Let $S = \{\sin x, \cos x\}$. This is a linearly independent set since

$$a_1 \sin x + a_2 \cos x = \mathbf{0} \implies f = \mathbf{0},$$

where $f(x) := a_1 \sin x + a_2 \cos x$ is the zero function, i.e., $f(x) = 0$ for all x. Choosing $x = 0$ and $x = \pi/2$ as arguments of f, we get $a_2 = 0$ and $a_1 = 0$, so that S is indeed linearly independent.

Example 5.5 Let $\mathbf{v}_1 = (1, 2), \mathbf{v}_2 = (2, 1)$ and let $S = \{\mathbf{v}_1 \mathbf{v}_2\} \subset \mathbb{Z}_3^2$. Then S is a linearly dependent set in $V = Z_3^2$ since the zero vector can be written as a non-trivial linear combination:

$$\mathbf{v}_1 \oplus \mathbf{v}_2 = (1,2) \oplus (2,1) = (3,3) \text{mod } 3 = (0,0).$$

We end this section with a result that confirms the importance of linear independence.

Proposition 5.6 — Uniqueness of representation. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ denote a set of vectors. Let $\mathbf{v} \in \text{span}S$. Then $\mathbf{v}$ has a unique representation as a linear combination of vectors in S if and only if S is a linearly independent set.

Proof. Let

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots a_k\mathbf{v}_k.$$

$\boxed{\Longrightarrow}$: we are given that the above representation is unique, i.e., there are no other coefficients that can represent $\mathbf{v}$ as a linear combination of vectors in S. Assume now that S is not a linearly independent set. We show that this assumption yields a contradiction. By definition, linear dependence implies that the zero vector can be written as a non-trivial linear combination:

$$\mathbf{0} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots b_k\mathbf{v}_k,$$

for some scalars $b_1, b_2, \ldots, b_k$, not all zero. Taking the sum of the above two relations we obtain

$$\mathbf{v} = (a_1 + b_1)\mathbf{v}_1 + (a_2 + b_2)\mathbf{v}_2 + \cdots + (a_k + b_k)\mathbf{v}_k =: c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k,$$

where $c_i := a_i + b_i, i = 1, \ldots, k$ satisfy $c_i - a_i = b_i$ are not all zero. Hence, $c_i \neq a_i$ for some i. Thus, the representation of $\mathbf{v}$ is not unique – a contradiction. Hence, S must be a linearly independent set.
$\boxed{\Longleftarrow}$: we are given that S is a linearly independent set. Assume that the coefficients $a_1, a_2, \ldots, a_k$ are not unique in the representation of $\mathbf{v}$. Then

$$\mathbf{v} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots b_k\mathbf{v}_k,$$

for some other coefficients $b_i \neq a_i$ for at least one i. Taking the difference, we find

$$\mathbf{0} = (a_1 - b_1)\mathbf{v}_1 + (a_2 - b_2)\mathbf{v}_2 + \cdots + (a_k - b_k)\mathbf{v}_k.$$

This implies that $\mathbf{0}$ can be written as a non-trivial linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$, which contradicts their assumed linear independence. Hence, we must have $a_i = b_i$ for all $i = 1, \ldots, k$ and the representation of $\mathbf{v}$ as a linear combination of $\mathbf{v}_i$ is unique. ■

The above result is an important characterisation of uniqueness of representation of a vector which highlights the importance of linear independence of sets. The following two sections provide a further insight into the concept of linear independence.

5.3 Removing vectors. Minimal spanning sets.

Lemma 5.7 Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$. Then S is a linearly dependent set if and only if there exists $j \in \{1, 2, \ldots, k\}$ such that $\text{span}S = \text{span}S \setminus \{\mathbf{v}_j\}$.

Proof. $\boxed{\Longrightarrow}$: we are given that S is a linearly dependent set; we need to show that there exists an index j such that $\text{span}S = \text{span}S \setminus \{\mathbf{v}_j\}$. To prove this set equivalence, we show that double inclusion holds. First, for any $j = 1, 2, \ldots, k$,

$$S \setminus \{\mathbf{v}_j\} \subseteq S \Longrightarrow \text{span}S \setminus \{\mathbf{v}_j\} \subseteq \text{span}S.$$

To show the second inclusion, let $\mathbf{v} \in \text{span}S$ have the representation

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k.$$

Since S is a linearly dependent set, the zero vector can be written as a non-trivial linear combination of vectors in S:

$$\mathbf{0} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_k\mathbf{v}_k.$$

This means that there exists an index j such that $b_j \neq 0$. Define $c = a_j/b_j$. Then

$$\text{span}S \ni \mathbf{v} = \mathbf{v} - c\mathbf{0} = (a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k) - c\,(b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_k\mathbf{v}_k) =$$

$$= \sum_{i=1}^{k}\left(a_i - \frac{a_j}{b_j}b_i\right)\mathbf{v}_i = \sum_{i\neq j}^{k}\left(a_i - \frac{a_j}{b_j}b_i\right)\mathbf{v}_i \in \text{span}S \setminus \{\mathbf{v}_j\}.$$

Since $\mathbf{v}$ was arbitrary, we conclude that $\text{span}S \subseteq \text{span}S \setminus \{\mathbf{v}_j\}$. Hence, $\text{span}S = \text{span}S \setminus \{\mathbf{v}_j\}$.
$\boxed{\Longleftarrow}$: we are given that $\text{span}S = \text{span}S \setminus \{\mathbf{v}_j\}$ for some j. To show S is linearly dependent, note that $\mathbf{v}_j \in \text{span}S = \text{span}S \setminus \{\mathbf{v}_j\}$. Thus, $\mathbf{v}_j \in \text{span}S \setminus \{\mathbf{v}_j\}$ so that

$$\mathbf{v}_j = \sum_{i\neq j} a_i\mathbf{v}_i \Longleftrightarrow a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + (-1)\mathbf{v}_j + \cdots + a_k\mathbf{v}_k = \mathbf{0},$$

which means that $\mathbf{0}$ is a non-trivial linear combination of the vectors in S (since at least the coefficient of $\mathbf{v}_j$ is nonzero). Hence, S is linearly dependent. ■

We illustrate this result with an example.

Example 5.6 Consider again the following vectors in $\mathbb{R}^3$:

$$\mathbf{v}_1 = \begin{bmatrix}1\\2\\0\end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix}1\\1\\0\end{bmatrix}.$$

The set $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent, so that there is a vector that we can remove without changing the span. In this particular case, any vector can be removed (i.e., the index j in Lemma 5.7 can be 1,2 or 3). However, one can notice that removing any vector yields a linearly independent set, so we cannot remove a further vector without changing the span.

Lemma 5.7 indicates that we can 'trim down' a linearly dependent set without changing its span: when this is the case, the representation of a vector will use fewer vectors from the span and so it will be more 'economical'. However, when the span changes, then we can identify a spanning set that is linearly independent. By definition, this is a minimal spanning set. The next result indicates that linear independence provides a characterisation of minimal spanning sets.

Theorem 5.8 Let S be a spanning set for a finite-dimensional vector space $V(\mathbb{F})$. Then S is a minimal spanning set for V if and only if it is a linearly independent set in V.

Proof. If $V(\mathbb{F})$ is the trivial vector space, then its minimal spanning set is the empty set, which is a linearly independent spanning set for V (see Definition 5.4). Let us assume now that $V(\mathbb{F})$ is a non-trivial vector space. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a (non-empty) spanning set for V.
$\boxed{\Longrightarrow}$: Let S be a minimal spanning set. We show that S is linearly independent. We do this by contradiction: assume that S is linearly dependent. By Lemma 5.7, $\text{span}S = \text{span}S \setminus \{\mathbf{v}_j\}$ for some j. Hence, $\text{span}S = V = \text{span}S \setminus \{\mathbf{v}_j\}$ and therefore S is not a minimal spanning set, as it does not satisfy

Definition 4.5.
$\boxed{\Longleftarrow}$: Let S be a linearly independent spanning set. We show that S is a minimal spanning set for V. We do this by contradiction: assume that S is not a minimal spanning set. Then there exists an index j such that $\text{span}S \setminus \{\mathbf{v}_j\} = V$. But $V = \text{span}S$. Hence, there exists an index j such that $\text{span}S = \text{span}S \setminus \{\mathbf{v}_j\}$ and by Lemma 5.7 this implies that S is a linearly dependent set – a contradiction. ■

5.4 Adding vectors. Maximal linearly independent sets.

Lemma 5.7 describes the change in the span of a set S under the removal of a vector from S. What happens when we augment S with a given vector? The following result provides an answer.

Lemma 5.9 Let S be linearly independent and let $\mathbf{v}$ be given. Then $S \cup \{\mathbf{v}\}$ is linearly independent if and only if $\mathbf{v} \notin \text{span}S$.

Proof. $\boxed{\Longrightarrow}$: let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ and $S \cup \{\mathbf{v}\}$ be linearly independent sets. We show that $\mathbf{v} \notin \text{span}S$. We do this by contradiction: assume $\mathbf{v} \in \text{span}S$. Then

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k \Longleftrightarrow a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k + (-1)\mathbf{v} = \mathbf{0}.$$

Hence $\mathbf{0}$ is a non-trivial linear combination of the vectors in the set $S \cup \{\mathbf{v}\}$; therefore, by definition, this set is linearly dependent, which contradicts our initial assumption.
$\boxed{\Longleftarrow}$: let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be linearly independent and assume $\mathbf{v} \notin \text{span}S$. We show that $S \cup \{\mathbf{v}\}$ is linearly independent. We do this again by contradiction: assume that $S \cup \{\mathbf{v}\}$ is linearly dependent. Then

$$\mathbf{0} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k + b\mathbf{v},$$

where not all coefficients are zero. We consider two cases:
(i) $b = 0$: we get

$$\mathbf{0} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k,$$

which implies $a_i = 0$ for all i, since S is linearly independent. Therefore, $\mathbf{0}$ is a trivial linear combination of vectors in $S \cup \{\mathbf{v}\}$, which implies that this set is linearly independent – a contradiction.
(ii) $b \neq 0$: in this case we can write $\mathbf{v}$ as follows:

$$\mathbf{v} = \frac{a_1}{b}\mathbf{v}_1 + \frac{a_2}{b}\mathbf{v}_2 + \cdots + \frac{a_k}{b}\mathbf{v}_k.$$

Hence $\mathbf{v} \in \text{span}S$ – a contradiction.
We conclude that $S \cup \{\mathbf{v}\}$ must be linearly independent. ■

Let us consider some examples.

Example 5.7 Consider the following vectors in $\mathbb{R}^3$:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Then $S = \{\mathbf{v}_1, \mathbf{v}_2\}$ is a linearly independent set and $\mathbf{v} \notin \text{span}S$. By Lemma 5.9, the set $S \cup \{\mathbf{v}\}$ is linearly independent, a fact which we can verify directly.

Example 5.8 Let $p_j(x) = x^j$, for $j = 0, 1, \ldots, k$. Let $S = \{p_0, p_1, \ldots, p_k\}$. Then S is linearly independent; one can immediately see that $\text{span}S = \mathcal{P}_k$, the space of polynomials of degree k. We can augment S

with p_{k+1}, which is not in span$S$; this results in the set $S \cup \{p_{k+1}\}$ which spans $\mathcal{P}_{k+1}$. Thus, polynomial spaces can be constructed by augmenting existing sets of monomials with monomials of higher degree.

Lemma 5.9 allows us to construct linearly independent sets of ever larger cardinality, by choosing new vectors in a suitable way. We can therefore augment successively linearly independent sets, while preserving the property of independence. However, this cannot be done indefinitely as otherwise V would have to be an infinite-dimensional vector space, which we excluded from our discussion. This observation suggests the following definition (compare it with Definition 4.5).

Definition 5.5 — Maximal linearly independent set. Let S be a linearly independent set of vectors from a vector space $V(\mathbb{F})$.

i. If $V(\mathbb{F})$ is trivial, the maximal linearly independent set if defined to be the empty set.
ii. If $V(\mathbb{F})$ is non-trivial, S is called a maximal linearly independent set in V if

$$S \cup \{\mathbf{v}\} \text{ is linearly dependent for any } \mathbf{v} \in V \setminus \{\mathbf{0}\}.$$

R The above definition considers separately the case where $V(\mathbb{F})$ is the trivial vector space, as there are no non-zero vectors to augment S with (so the criterion for the non-trivial case ii. does not make sense). Note that the definition 'works' if $V(\mathbb{F})$ is the trivial vector space: by Proposition 5.2, $S = \emptyset$ is linearly independent.

We end with the following characterisation of maximal linearly independent sets.

Theorem 5.10 Let S be a linearly independent set of vectors from a finite-dimensional vector space $V(\mathbb{F})$. Then S is a maximal linearly indepedent set in V if and only it is a spanning set for V.

Proof. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a linearly independent set in V.
$\boxed{\Longrightarrow}$ Let S be maximal. Let $\mathbf{v} \in V \setminus \{\mathbf{0}\}$. By the definition of maximal set, the set $S \cup \{\mathbf{v}\}$ is linearly dependent. Hence, there exist scalars $a_1, a_2, \ldots, a_k$, not all zero, such that

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k.$$

Since $\mathbf{v} \in V \setminus \{\mathbf{0}\}$ is arbitrary, we conclude that $V \setminus \{\mathbf{0}\} \subseteq \text{span}S$. Given that $\mathbf{0} \in \text{span}S$, we conclude that $V \subseteq \text{span}S$. On the other hand, span$S \subseteq V$ (by closure). Hence, $V = \text{span}S$ and therefore S is a spanning set for V.
$\boxed{\Longleftarrow}$ Let S be a linearly indepedent spanning set. We need to show that S is maximal. By Theorem 5.8, S is a minimal spanning set for V and therefore any $\mathbf{v} \in V \setminus \{\mathbf{0}\}$ can be written as a non-trivial linear combination

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k \Longleftrightarrow \mathbf{0} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k - \mathbf{v},$$

which means that the zero vector can be written as a non-trivial linear combination of vectors in the set $S \cup \{\mathbf{v}\}$. Therefore, this set is linearly dependent. Hence, by Definition 5.5, S is a maximal linear independent set in V. ■

The above discussion indicates that minimal spanning sets and maximal linearly independent sets are special and should be further investigated. This is the topic of the next lecture.

LECTURE 6

Bases. Dimension

6.1 Basis sets

The discussion in the previous lecture suggests that minimal spanning sets and maximal linearly independent sets are the same. This is indeed the case.

Theorem 6.1 Let S be a set of vectors in a finite-dimensional vector space $V(\mathbb{F})$. Then S is a minimal spanning set for V if and only if it is a maximal linearly independent set in V.

Proof. The result follows from the characterisations provided by Theorems 5.8 and 5.10. ■

These sets are also known as basis sets. A common definition is included below.

Definition 6.1 — Basis. A basis set for a vector space $V(\mathbb{F})$ is a set S satisfying

- S is a spanning set for $V(\mathbb{F})$;
- S is a linearly independent set.

To establish if a given set is a basis for a vector space, we need to check both the spanning property and the linear independence of the set.

Example 6.1 — Canonical basis for $\mathbb{R}^3$. Let

$$B = \left\{ \mathbf{e}_1 := \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e}_2 := \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{e}_3 := \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

It is straightforward to see that B spans $\mathbb{R}^3$ and is a linearly independent set, hence it is a basis for $\mathbb{R}^3$, known as the canonical basis (with the vectors $\mathbf{e}_i$ known as canonical vectors). The generalisation to $\mathbb{R}^n$ is straightforward.

Example 6.2 — Power basis for $\mathcal{P}_2$. Let $B = \{1, x, x^2\}$. Then B is a basis for $\mathcal{P}_2$, known as the power basis. The generalisation to $\mathcal{P}_n$ is straightforward.

By Theorem 5.8, a basis set is a minimal spanning set for $V(\mathbb{F})$. By the discussion in the previous section, we deduce that

- basis sets always exist;
- basis sets are not unique.

The non-uniqueness of a basis set suggests that some bases may be preferred over others: this is indeed the case and we will see later how to identify and construct bases that are convenient in a given context.

The following results describe further properties of basis sets.

Proposition 6.2 Let $V(\mathbb{F})$ be a vector space. Every spanning set for V contains a basis for V.

Proof. Since any basis is a minimal spanning set, the result follows from Corollary 4.4. ■

Proposition 6.3 Let $V(\mathbb{F})$ be a vector space. Every linearly independent set in V is contained in some basis set for V.

Proof. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ be a linearly independent set in $V = \text{span}U$, where $U = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_\ell\}$. We consider the following two complementing cases.
1. Assume that $\mathbf{u}_i \in \text{span}S$ for all $i = 1, 2, \dots, \ell$. Then, $V = \text{span}U \subseteq \text{span}S$. However, since $\mathbf{v}_i \in V$, we have $\text{span}S \subseteq V = \text{span}U$. Therefore $\text{span}S = \text{span}U = V$. This means that S is a linearly independent spanning set for V and therefore a basis. In this case, the result holds trivially since the set S is contained in the basis S.
2. Assume now that $\mathbf{u}_i \notin \text{span}S$ for some i. Then, by Lemma 5.9, $S' := S \cup \{\mathbf{u}_i\}$ is a linearly independent set. If S' is a spanning set for V, then it is a basis and $S \subset S'$, which proves the statement. Otherwise, we replace S with S' and repeat the reasoning in 1 and 2 to obtain a new linearly independent set S' which contains the original set S. If S' is a spanning set, then it is the basis we seek. If not, we continue this procedure: we can do this at most ℓ times; if this indeed happens, we obtain $S' = S \cup U$, in which case S' is a spanning set because U is one. Therefore S' is a basis with $S \subset S'$. ■

The above proof can be used to derive the following corollary.

Corollary 6.4 Let $V(\mathbb{F})$ be a vector space. Let S be a linearly independent set in V and let U be a spanning set for V. Then $|S| \le |U|$.

These results allow us to derive the following important fact.

Theorem 6.5 Any two basis sets for a finite-dimensional vector space have the same number of elements.

Proof. Let B_1, B_2 be two basis sets for a given vector space. By definition, they are linearly independent spanning sets. This allows us to set $B_1 = U, B_2 = S$ in Corollary 6.4; we deduce that $|B_2| \le |B_1|$. Reversing the roles of B_1 and B_2, we find $|B_1| \le |B_2|$, so that $|B_1| = |B_2|$, as stated in the theorem. ■

6.2 Dimension

The result in the previous theorem allows for the following definition.

Definition 6.2 — Dimension. Let $V(\mathbb{F})$ be a non-trivial finite-dimensional vector space. The dimension of V, denoted by $\dim V$, is the number of vectors in a basis for V. If V is trivial, then $\dim V = 0$.

The following results are fairly intuitive.

Proposition 6.6 Let $U \leq V$. Then $\dim U \leq \dim V$, with $\dim U = \dim V$ if and only if $U = V$.

Proof. Let B_U, B_V be basis sets for U and V, respectively. Then B_U is a linearly independent set in V and B_V is a spanning set for V. By Corollary 6.4, $|B_U| \leq |B_V|$ and therefore $\dim U \leq \dim V$.
To show the last statement, let B_U be a basis for U. By Proposition 6.3, we have $B_U \subset B_V$ for some basis set for V. Since, by hypothesis, $\dim U = \dim V$, we have $|B_U| = |B_V|$, so that $B_U = B_V$. Then $U = \text{span}B_U = \text{span}B_V = V$.
On the other hand, if $U = V$, then $U \leq V$ and also $V \leq U$. Hence, by the first statement of the proposition, $\dim U \leq \dim V$ and also $\dim V \leq \dim U$, so that $\dim U = \dim V$. ■

Proposition 6.7 Let $V(\mathbb{F})$ be a non-trivial finite-dimensional vector space. Let S be a linearly independent set in V with $|S| = \dim V$. Then S is a basis set for V.

Proof. Since S is linearly independent, it is a basis for span$S$, which is a subspace for V: $\text{span}S \leq V$. Since $|S| = \dim V$, by Proposition 6.6, $V = \text{span}S$. Hence S is a spanning set for V, which is also linearly independent, so it is a basis, by definition. ■

We end this section with the following important result.

Theorem 6.8 Let U, V be two subspaces of a finite-dimensional vector space. Then

$$\dim\,(U + V) = \dim U + \dim V - \dim U \cap V.$$

Proof. Let $W := U \cap V$ and define

$$\dim U =: k, \quad \dim V =: \ell, \quad \dim W = m.$$

Let B_W be a basis for W. By Proposition 6.3, since B_W is a linearly independent set in U, it must be contained in some basis B_U of U. Similarly, B_W must be contained in some basis B_V of V. Let us denote these bases as follows:

- $B_W = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m\}$,
- $B_U := \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m, \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{k-m}\}$,
- $B_V := \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{\ell-m}\}$,

where the basis sets B_U, B_V contain elements $\mathbf{u}_j \notin V$ and $\mathbf{v}_s \notin U$, respectively. Then a spanning set for $U + V$ is

$$B := \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m, \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{k-m}, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{\ell-m}\}.$$

This can be readily seen since every element in $U + V$ is a sum of an element of U (hence a linear combination of elements in B_U and an element of V (hence a linear combination of elements in B_V). If B is a linearly independent set, it is a basis set for $U + V$. Therefore,

$$\dim\,(U + V) = |B| = m + (k - m) + (\ell - m) = k + \ell - m = \dim U + \dim V - \dim U \cap V.$$

It remains to show B is indeed linearly independent. Consider the representation of $\mathbf{0}$ as a linear combinations of elements of B:

$$\mathbf{0} = \sum_{i=1}^{m} a_i \mathbf{w}_i + \sum_{j=1}^{k-m} b_j \mathbf{u}_j + \sum_{s=1}^{\ell-m} c_s \mathbf{v}_s.$$

We can re-write this as

$$\sum_{i=1}^{m} a_i \mathbf{w}_i + \sum_{j=1}^{k-m} b_j \mathbf{u}_j = -\sum_{s=1}^{\ell-m} c_s \mathbf{v}_s.$$

The expression on the left represents a vector in U, while that on the right a vector in V. The equality implies that they both represent a vector in both U and V, i.e., in $W = U \cap V$. Since any vector in W is uniquely expressed as a linear combination of $\mathbf{w}_i$, we must have $b_j = 0$ for all j and also $c_s = 0$ for all s. However, this results in another representation of the zero vector involving the elements $\mathbf{w}_i$ of B_W:

$$\sum_{i=1}^{m} a_i \mathbf{w}_i = \mathbf{0}.$$

Since B_W is a linearly independent set, it follows that $a_i =$ for all i. Thus, the intial representation of $\mathbf{0}$ is trivial and the set B is linearly independent. ■

An immediate consequence of this result can be established for the case of direct sums of vector spaces.

Corollary 6.9 Let U, V be two subspaces of a finite-dimensional vector space and define $X = U + V$. Then

$$X = U \oplus V \iff \dim X = \dim U + \dim V.$$

Proof. Let $X = U \oplus V$. By the direct sum criterion 1 we have $U \cap V = \{\mathbf{0}\}$. Hence, since $\dim \operatorname{span}\{\mathbf{0}\} = 0$, we get

$$\dim X = \dim U + \dim V.$$

Conversely, let $X = U + V$ and assume $\dim X = \dim U + \dim V$. By Theorem 6.8, we must have $\dim U \cap V = 0$. Hence, $U \cap V = \{\mathbf{0}\}$ and therefore $X = U \oplus V$. ■

6.3 Coordinates

Recall that given a linearly independent set, we can represent uniquely any vector in its span (see Proposition 5.6). This is an important property which we can employ in the case when the set is a basis for some vector space V. In particular, given a basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ for a vector space $V(\mathbb{F})$, there exist unique coefficients $x_i \in \mathbb{F}$ such that any $\mathbf{v} \in V(\mathbb{F})$ can be written as

$$\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \ldots + x_n\mathbf{v}_n.$$

Given their uniqueness, we can provide the following definition.

Definition 6.3 The coefficients $x_1, x_2, \ldots, x_n$ are called the **coordinates** of the vector $\mathbf{v}$ in the basis B.

Note that the coordinates of a vector depend on the choice of basis, which is why sometimes they are also referred to as B-coordinates.

Proposition 6.10 — Coordinate map. Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ denote a basis for V. Let

$$\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \ldots + x_n\mathbf{v}_n.$$

Define the **coordinate map** $\varphi : V \mapsto \mathbb{F}^n$ via

$$\varphi(\mathbf{v}) = \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ v_n \end{bmatrix}.$$

Then φ is a bijection that satisfies the linearity condition

$$\varphi(a\mathbf{x} + b\mathbf{y}) = a\varphi(\mathbf{x}) + b\varphi(\mathbf{y}).$$

R This is an example of a so-called **isomorphism**. We will investigate this topic in Part II when we discuss linear maps. For now, we recall that a bijection, as a one-to-one correspondence, has a unique inverse $\varphi^{-1} : \mathbb{F}^n \mapsto V$: thus, if $\varphi(\mathbf{v}) = \mathbf{x}$, we also have $\varphi^{-1}(\mathbf{x}) = \mathbf{v}$.

Depending on the application of interest, a choice of basis may be preferred over others: often, this choice relates to the resulting coordinates or other evaluations involving basis elements. A suitable basis could yield coordinates that have some physical significance, and/or a sparse set of coordinates, i.e., with only few non-zero values. We will encounter some examples later; for now, we discuss two examples of standard bases.

Example 6.3 — Coordinates in the canonical basis. Let $\mathbb{R}^3$ denote Euclidean space. Recall that the canonical basis was defined to be

$$B = \left\{ \mathbf{e}_1 := \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},\ \mathbf{e}_2 := \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},\ \mathbf{e}_3 := \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Then any vector in $\mathbb{R}^3$ can be represented in the following convenient form

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} v_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ v_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ v_3 \end{bmatrix} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3.$$

Thus, the coordinates in the canonical basis are simply the entries in the vector $\mathbf{v}$. In Euclidean space, these coordinates have the usual (geometric) significance.

Example 6.4 Let $B = \left\{1, x, x^2\right\}$ denote the power basis for $\mathcal{P}_2$. Then any polynomial in $\mathcal{P}_2$ can be represented in the form

$$p(x) = a_0 + a_1 x + a_2 x^2,$$

so that the coordinates in the power basis are just the polynomial coefficients.

We end with an example where a non-standard basis is employed.

Example 6.5 Let

$$B' = \left\{ \mathbf{v}_1 := \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},\ \mathbf{v}_2 := \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix},\ \mathbf{v}_3 := \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Let $\mathbf{v} \in \mathbb{R}^3$ be a generic vector and let us compute its coordinates in the basis B':

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 = \begin{bmatrix} a_1 \\ a_1 + a_2 \\ a_1 + a_2 + a_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \Longrightarrow \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 - v_1 \\ v_3 - v_2 \end{bmatrix}.$$

Exercise 6.1 Let $B' = \{p_1, p_2, p_3\}$, where $p_i = 1 - (i-1)x^{i-1}$ for $i = 1, 2, 3$. Find the coordinates of the polynomial $p(x) = 1 + 2x + 3x^2$ in the basis B'.

LECTURE 7

Inner product spaces (1)

In Lecture 1, we reviewed several operations involving vectors. One of them was the dot product, which allowed for the concepts of angle between vectors, length and orthogonality. In this lecture, we will generalise the concept of dot product to vectors other than Euclidean. Vector spaces equipped with this type of operation will afford additional structure by extending some familiar concepts of Euclidean geometry.

7.1 Inner products

The dot (or scalar) product is an operation between two vectors which returns a scalar. In the case of Euclidean vectors, this scalar was a real number. We recall it below:

$$\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}|\,|\mathbf{B}| \cos \alpha,$$

where α is the angle between the direction lines of $\mathbf{A}$ and $\mathbf{B}$.

It is not immediately clear how this definition can be extended to general vector spaces V: what is the angle between two vectors in $\mathbb{R}^4$ or that between two polynomials in $\mathcal{P}_n$? Instead of using the above definition in its form, we aim to extract some of the features it encapsulates.

We note the following three properties associated with the dot product:

- commutativty (or symmetry): $\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A}$;
- non-negativity: $\mathbf{A} \cdot \mathbf{A} \geq 0$, with $\mathbf{A} \cdot \mathbf{A} = 0$ if and only if $\mathbf{A} = \mathbf{0}$;
- linearity: $(a\mathbf{A} + b\mathbf{B}) \cdot \mathbf{C} = a\mathbf{A} \cdot \mathbf{C} + b\mathbf{B} \cdot \mathbf{C}$.

Note that while the first two can be immediately verified, the third is best checked using the representations of $\mathbf{A}, \mathbf{B}, \mathbf{C}$ in the orthonormal basis $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$.

We extend the definition of this operation to the case of a real finite-dimensional vector space by using these three properties. Before we give the definition, we note that the dot product can be viewed as a function of two arguments which returns real values:

$$\mathcal{B}(\mathbf{A}, \mathbf{B}) = |\mathbf{A}|\,|\mathbf{B}| \cos \alpha.$$

Moreover, we note that $\mathcal{B}(\cdot, \cdot)$ is linear in each argument. These observations are formalised in the following definition.

Definition 7.1 Let U, V, W denote vector spaces. A **bilinear function** is a function of two arguments $\mathcal{B}(\cdot,\cdot) : V \times W \to U$, satisfying the following linearity properties.
- $\mathcal{B}(a\mathbf{u} + b\mathbf{v}, \mathbf{w}) = a\mathcal{B}(\mathbf{u}, \mathbf{w}) + b\mathcal{B}(\mathbf{v}, \mathbf{w})$,
- $\mathcal{B}(\mathbf{u}, a\mathbf{v} + b\mathbf{w}) = a\mathcal{B}(\mathbf{u}, \mathbf{v}) + b\mathcal{B}(\mathbf{u}, \mathbf{w})$.

We say $\mathcal{B}(\cdot,\cdot)$ is **symmetric** if $\mathcal{B}(\mathbf{v}, \mathbf{w}) = \mathcal{B}(\mathbf{w}, \mathbf{v})$.

In the case of the dot product, one can view the domain of $\mathcal{B}(\cdot,\cdot)$ as $\mathbb{E}^3 \times \mathbb{E}^3$, with the codomain being $\mathbb{R}$: $\mathcal{B}\colon \mathbb{E}^3 \times \mathbb{E}^3 \to \mathbb{R}$. In the definition below, we replace the generic notation $\mathcal{B}(\cdot,\cdot)$ with $\langle\cdot,\cdot\rangle$.

Definition 7.2 — Inner product. Let $V(\mathbb{R})$ be a real vector space. A real inner product is a function $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ which satisfies the following properties:

i. symmetry: $\langle \mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{v}, \mathbf{u}\rangle$;
ii. linearity: $\langle a\mathbf{u} + b\mathbf{v}, \mathbf{w}\rangle = a\langle \mathbf{u}, \mathbf{w}\rangle + b\langle \mathbf{v}, \mathbf{w}\rangle$;
iii. non-negativity: $\langle \mathbf{v}, \mathbf{v}\rangle \geq 0$, with $\langle \mathbf{v}, \mathbf{v}\rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

Property ii. requires linearity in the first argument of the inner product. However, by symmetry, we have linearity also in the second argument, as the following exercise shows.

Exercise 7.1 Let V be a real vector space equipped with a real inner product. Show that for any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $a, b \in \mathbb{R}$, properties i. and ii. imply that $\langle\cdot,\cdot\rangle$ is linear in its second argument:

$$\langle \mathbf{u}, a\mathbf{v} + b\mathbf{w}\rangle = a\langle \mathbf{u}, \mathbf{v}\rangle + b\langle \mathbf{u}, \mathbf{w}\rangle.$$

R Due to linearity in both arguments, inner products are referred to as **bilinear functions** or **bilinear forms**. Thus, an inner product is a symmetric bilinear form which is zero only at the zero vector.

Example 7.1 We define the **standard/Euclidean inner product** on $\mathbb{R}^n$ via

$$\langle \mathbf{u}, \mathbf{v}\rangle = u_1v_1 + u_2v_2 + \cdots + u_nv_n.$$

It is straightforward to verify the three properties in the definition.

Example 7.2 We define the inner product on $\mathcal{P}_n([a,b])$ via

$$\langle p, q\rangle = \int_a^b p(x)q(x)\mathrm{d}x.$$

Properties i. and ii. are straightforward to verify; property iii. is verified using the standard properties of integrals.

R Note that the notation $\mathcal{P}_n([a,b])$ indicates the set of polynomials defined on the closed interval $[a,b]$. This choice ensures that the integral is a real number, which is an implicit requirement in the definition of the inner product.

The previous example indicates that property iii. may be the usual focus when establishing when a bilinear function is an inner product. Indeed, here is an example where it fails to be satisfied.

Example 7.3 Let $V = \mathcal{P}_n([a, b])$ and define $\mathcal{F} : V \times V \to \mathbb{R}$ via

$$\mathcal{F}(p, q) = \int_a^b p'(x)q'(x)\mathrm{d}x.$$

While $\mathcal{F}$ is a non-negative symmetric bilinear form, it fails to be zero only at the zero vector since the choice $p(x) = 1$ yields

$$\mathcal{F}(p, p) = \int_a^b [p'(x)]^2 \mathrm{d}x = \int_a^b 0\mathrm{d}x = 0.$$

Hence, $\mathcal{F}$ is not an inner product.

Exercise 7.2 Let $V = C^1([a, b])$ and consider the following candidate for an inner product:

$$\mathcal{F}(f, g) = \int_a^b \left[f(x)g(x) + f'(x)g'(x)\right] \mathrm{d}x.$$

Show that $\mathcal{F}$ is an inner product.

Definition 7.2 is restricted to the case of vector spaces over the reals. Can we have fields $\mathbb{F}$ other than the reals, i.e., can we have $\langle\cdot,\cdot\rangle : V(\mathbb{F}) \times V(\mathbb{F}) \to \mathbb{F}$? The answer depends on the validity of property iii. The inequality $\langle\mathbf{v},\mathbf{v}\rangle \geq 0$ implies that $\langle\mathbf{v},\mathbf{v}\rangle$ takes a value in an ordered field. This cannot be $\mathbb{C}$, nor any finite field. However, if we make the choice $\mathbb{F} = \mathbb{C}$, we can redefine property i., so that $\langle\mathbf{v},\mathbf{v}\rangle$ is a real number and property iii. makes sense.

Definition 7.3 — Complex inner product. Let $V(\mathbb{C})$ be a complex vector space. A complex inner product is a function $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{C}$ which satisfies the following properties:

i. conjugate symmetry: $\langle\mathbf{u},\mathbf{v}\rangle = \overline{\langle\mathbf{v},\mathbf{u}\rangle}$;
ii. linearity: $\langle a\mathbf{u} + b\mathbf{v},\mathbf{w}\rangle = a\langle\mathbf{u},\mathbf{w}\rangle + b\langle\mathbf{v},\mathbf{w}\rangle$;
iii. non-negativity: $\langle\mathbf{v},\mathbf{v}\rangle \geq 0$, with $\langle\mathbf{v},\mathbf{v}\rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

Exercise 7.3 Let $V(\mathbb{C})$ be a vector space equipped with a complex inner product. Show that for any vector $\mathbf{v} \in V$ there holds $\langle\mathbf{v},\mathbf{v}\rangle \in \mathbb{R}$.

In the following, we will focus our attention mostly on real vector spaces, although we will occasionally comment also on the complex case.

7.2 Inner product spaces

Inner products provide additional structure to vector spaces. We will focus in the remainder of this course on vector spaces equipped with a real inner product.

Definition 7.4 — Inner product spaces. A vector space $V(\mathbb{F})$ equipped with an inner product $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{F}$ is called an inner product space.

To indicate that $V(\mathbb{F})$ is an inner product space, we sometimes write $(V(\mathbb{F}), \langle\cdot,\cdot\rangle)$ or $(V, \langle\cdot,\cdot\rangle)$, although often the inner product will be evident from the context.

R One can define more than one inner product on a vector space. The choice usually is provided and/or justified by applications.

Given an inner product space, we can immediately introduce two concepts analogous to those used for Euclidean spaces, namely, length (or norm) of a vector and angle between vectors.

7.2.1 Length

Definition 7.5 — Length/norm of a vector. Let $V(\mathbb{F})$ be an inner product space equipped with an inner product $\langle \cdot, \cdot \rangle$. For any vector $\mathbf{v} \in V$ we denote its length or norm by $\|\mathbf{v}\|$ and define it via

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}.$$

Note that the definition makes sense (i.e., the length is defined to be a positive real number, unless the vector is the zero vector) due to property iii. of inner products. This means that every inner product space has automatically a norm function defined on it: $\|\cdot\| : V \to \mathbb{R}_+ \cup \{0\}$. This makes V a **normed vector space**; the norm is called the **induced norm**. To indicate this relationship, the inner-product on V and the corresponding induced norm are sometimes denoted by $\langle \cdot, \cdot \rangle_V$, $\|\cdot\|_V$, respectively.
This definition allows us to establish the following results and properties involving norms:
- Cauchy-Schwarz inequality;
- triangle inequality;
- length of a scaled vector.

We derive each result below.

Proposition 7.1 — Cauchy-Schwarz inequality. Let $\|\cdot\|$ denote the norm induced by a real inner product on a vector space $V(\mathbb{R})$. Then

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \le \|\mathbf{u}\| \, \|\mathbf{v}\| \quad \text{for any } \mathbf{u}, \mathbf{v} \in V.$$

Proof. Let $\mathbf{u}, \mathbf{v} \in V$ and let $a \in \mathbb{R}$. Then, using the properties of inner-products,

$$0 \le \|\mathbf{u} + a\mathbf{v}\|^2 = \langle \mathbf{u} + a\mathbf{v}, \mathbf{u} + a\mathbf{v} \rangle = \langle \mathbf{u}, u \rangle + 2a \langle \mathbf{u}, \mathbf{v} \rangle + a^2 \langle \mathbf{v}, \mathbf{v} \rangle = \|\mathbf{u}\|^2 + 2a \langle \mathbf{u}, \mathbf{v} \rangle + a^2 \|\mathbf{v}\|^2 .$$

This is a quadratic inequality in a, which holds provided the discriminant associated with the quadratic on the right is non-positive:

$$\Delta := 4 \langle \mathbf{u}, \mathbf{v} \rangle^2 - 4 \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \le 0 \Longrightarrow |\langle \mathbf{u}, \mathbf{v} \rangle| \le \|\mathbf{u}\| \, \|\mathbf{v}\| .$$

■

The triangle inequality can be viewed as corollary of the Cauchy-Schwarz inequality.

Proposition 7.2 — Triangle inequality. Let $V(\mathbb{R})$ be an inner product space. Then for any $\mathbf{u}, \mathbf{v} \in V$ there holds

$$\|\mathbf{u} + \mathbf{v}\| \le \|\mathbf{u}\| + \|\mathbf{v}\| .$$

Proof. We have, using the Cauchy-Schwarz inequality,

$$\|\mathbf{u} + \mathbf{v}\|^2 = \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle = \|\mathbf{u}\|^2 + 2 \langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2 \le \|\mathbf{u}\|^2 + 2 \|\mathbf{u}\| \, \|\mathbf{v}\| + \|\mathbf{v}\|^2 = (\|\mathbf{u}\| + \|\mathbf{v}\|)^2$$

and the result follows by taking square-roots on both sides. ■

Proposition 7.3 — Length of scaled vector. Let $V(\mathbb{R})$ be an inner product space. Then for any $\mathbf{v} \in V$ and $a \in \mathbb{R}$, there holds

$$\|a\mathbf{v}\| = |a| \, \|\mathbf{v}\| .$$

Proof. We have, using the properties of the inner product,

$$\|a\mathbf{v}\|^2 = \langle a\mathbf{v}, a\mathbf{v}\rangle = a^2 \langle \mathbf{v}, \mathbf{v}\rangle = a^2 \|\mathbf{v}\|^2 .$$

The result follows by taking square-roots on both sides. ■

R This result confirms that multiplying a vector by a general scalar a results in a vector with length multiplied by $|a|$. Indeed, the sign of a should not (and does not) play a role in the resulting length.

Before we consider some examples, we note that given any vector $\mathbf{v}$ in an normed space, we can always associate with it a **unit vector** defined via

$$\widehat{\mathbf{v}} := \frac{\mathbf{v}}{\|\mathbf{v}\|}.$$

This is simply a scaled vector with $a = 1/\|\mathbf{v}\|$, so that $\|\widehat{\mathbf{v}}\| = 1$.

Let us now consider some examples.

Example 7.4 — Cauchy-Schwarz inequality on $\mathbb{R}^n$. Let $\langle\cdot,\cdot\rangle$ denote the standard (Euclidean) inner product on $\mathbb{R}^n$. Then, by the Cauchy-Schwarz inequality, for any vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ there holds

$$\left|\sum_{i=1}^{n} u_i v_i\right| \leq \left(\sum_{i=1}^{n} u_i^2\right)^{1/2} \left(\sum_{i=1}^{n} v_i^2\right)^{1/2}.$$

Example 7.5 — Cauchy-Schwarz inequality on $C^0([a,b])$. Let $\langle\cdot,\cdot\rangle$ be defined via

$$\langle f, g\rangle = \int_a^b f(x)g(x)\mathrm{d}x.$$

Then, by the Cauchy-Schwarz inequality, for any continuous functions defined on $[a,b]$, there holds

$$\int_a^b f(x)g(x)\mathrm{d}x \leq \left(\int_a^b f(x)^2\mathrm{d}x\right)^{1/2} \left(\int_a^b g(x)^2\mathrm{d}x\right)^{1/2}.$$

7.2.2 Angles

The definition of angle between vectors can be extended directly from the Euclidean case to general inner product spaces

Definition 7.6 — Angle between vectors. Let $V(\mathbb{F})$ be an inner product space equipped with an inner product $\langle\cdot,\cdot\rangle$. The angle α between two non-zero vectors $\mathbf{u}, \mathbf{v} \in V$ is defined via

$$\cos\alpha = \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|}.$$

R The Cauchy-Schwarz inequality implies that, if $\mathbf{u}, \mathbf{v}$ are non-zero,

$$-\left|\langle \mathbf{u}, \mathbf{v}\rangle\right| \leq \langle \mathbf{u}, \mathbf{v}\rangle \leq \left|\langle \mathbf{u}, \mathbf{v}\rangle\right| \iff -1 \leq \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|} \leq 1.$$

This confirms that the definition of the angle between vectors $\mathbf{u}$ and $\mathbf{v}$ is well-posed (given that for any angle α there holds $-1 \leq \cos\alpha \leq 1$).

The concept of angle between vectors leads naturally to the concept of orthogonal vectors.

Definition 7.7 — Orthogonal vectors. Let $V(\mathbb{R})$ denote an inner product space. Then the nonzero vectors $\mathbf{u}, \mathbf{v} \in V$ are said to be orthogonal if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. We write $\mathbf{u} \perp \mathbf{v}$.

Orthogonal vectors in an inner product space satisfy the following characterisation.

Proposition 7.4 — Pythagoras. Let $V(\mathbb{R})$ denote an inner product space. Let $\mathbf{u}, \mathbf{v} \in V \setminus \{\mathbf{0}\}$. Then

$$\|\mathbf{u} \pm \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \iff \mathbf{u} \perp \mathbf{v}.$$

Proof. Follows from the identity (worth remembering)

$$\|\mathbf{u} \pm \mathbf{v}\|^2 = \|\mathbf{u}\|^2 \pm 2 \langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2 .$$

■

Example 7.6 Let $r, s \in \mathbb{N}$ and $a > 0$. Define $p(x) = x^{2r}, q(x) = x^{2s-1}$. Then p and q are orthogonal in the inner product

$$\langle p, q \rangle := \int_{-a}^{a} p(x)q(x)\mathrm{d}x.$$

LECTURE 8

Inner product spaces (2)

8.1 Orthogonal bases

Let V be an inner product space and let B denote a basis for V. Choosing the elements of B as orthogonal vectors has many benefits and is common in many applications. We consider such a choice below, including some examples.

Definition 8.1 — Orthogonal set. Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ be a set of nonzero vectors in V. Then S is said to be orthogonal if for all $i \neq j$

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0.$$

Note that, by definition, orthogonal sets do not include the zero vector. The advantage of working with such sets can be derived immediately.

Proposition 8.1 Orthogonal sets are linearly independent.

Proof. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ be an orthogonal set of nonzero vectors in an inner product space $(V, \langle \cdot, \cdot \rangle)$. Consider the linear combination

$$\mathbf{0} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k.$$

Then for any $j = 1, 2, \dots, k$,

$$0 = \langle \mathbf{0}, \mathbf{v}_j \rangle = \langle a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k, \mathbf{v}_j \rangle = a_j \left\| \mathbf{v}_j \right\|^2 \Longrightarrow a_j = 0.$$

Hence, the zero vector can only be written as the trivial linear combination of vectors in S and therefore S is linearly independent. ■

Let now V be an inner product space with dimension $\dim V = n$. Choosing a set of n orthogonal vectors results in a basis for V.

Definition 8.2 — Orthogonal basis. Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. A basis set is said to be an orthogonal basis for V if it is an orthogonal set.

We immediately derive the following property of orthogonal sets.

Proposition 8.2 Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space, with $\dim V = n$. Let S be an orthogonal set of vectors from V with $|S| = n$. Then S is an orthogonal basis for V.

Proof. By Proposition 8.1, S is linearly independent. The result follows by applying Proposition 6.7. ∎

Orthogonal bases allow for the coordinates of a generic vector $\mathbf{v} \in V$ to be computed via evaluations of inner products. This is a key advantage over computing the coordinates in the usual way, by solving a (possibly large) linear system of equations.

Proposition 8.3 Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ denote an orthogonal basis for V. Then the coordinates a_i of any vector $\mathbf{v} \in V$ are given by

$$a_i = \frac{\langle \mathbf{v}, \mathbf{v}_i \rangle}{\langle \mathbf{v}_i, \mathbf{v}_i \rangle}.$$

Proof. Let $\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots a_n\mathbf{v}_n$. Then, using orthogonality,

$$\langle \mathbf{v}, \mathbf{v}_i \rangle = \langle a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n, \mathbf{v}_i \rangle = a_i \langle \mathbf{v}_i, \mathbf{v}_i \rangle ,$$

and the result follows. ∎

R The coordinates a_i in Proposition 8.3 are known as the **Fourier coefficients** of $\mathbf{v}$ in the basis B.

When the elements of an orthogonal set are unit vectors, the manipulations and expressions arising are further simplified. In this case, we use the convention that the generic ith unit vector is denoted by $\mathbf{e}_i$.

Definition 8.3 — Orthonormal set. Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. A set S of nonzero vectors $\mathbf{e}_i$ in V is said to orthonormal if

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij} := \begin{cases} 0 & i \neq j, \\ 1 & i = j, \end{cases}$$

for all $i, j = 1, 2 \ldots, |S|$.

Correspondingly, we have the following definition.

Definition 8.4 — Orthonormal basis. Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. A set is said to be an orthonormal basis for V if

- it is an orthonormal set in V;
- it is a basis set for V.

The representation of the Fourier coefficents of a vector $\mathbf{v}$ is further simplified when working with an orthonormal basis:

$$a_i = \langle \mathbf{v}, \mathbf{e}_i \rangle .$$

In the context of coordinates, there is one more significant result concerning the evaluation of inner products.

Proposition 8.4 Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space with dimension n. Let $B = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ denote an orthonormal basis for V. Let $\mathbf{u}, \mathbf{v}$ have respective coordinates a_i, b_j, with $i, j = 1, \ldots, n$ Then

$$\langle \mathbf{u}, \mathbf{v} \rangle = a_1b_1 + a_2b_2 + \cdots a_nb_n.$$

Proof. We have, using orthogonality,

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \cdots a_n\mathbf{e}_n, b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + \cdots b_n\mathbf{e}_n \rangle = a_1b_1 + a_2b_2 + \cdots a_nb_n.$$

■

R The result of Proposition 8.4 is related to one of our previous observations, namely, that there is a one-to-one correspondence between any finite-dimensional vector space V and $\mathbb{R}^n$. In this case we witness the correspondence between an inner product of two vectors in V and the Euclidean inner product of their respective coordinates with respect to some orthonormal basis. This correspondence will be discussed in Part II.

8.2 Projections

The concept of orthogonality in a vector space allows for the extension of the geometric concept of orthogonal projection. In turn, orthogonal projections will allow us to decompose a vector into a sum of projections: a so-called orthogonal decomposition. These results will enable us to devise a procedure for constructing an orthonormal basis for any inner product space in the next lecture.

Let us start by aiming to write a vector as a sum of two vectors: one parallel to a fixed direction $\mathbf{u}$ and the other orthogonal to it. We have

$$\mathbf{v} = \mathbf{v}^{\parallel} + \mathbf{v}^{\perp},$$

where $\mathbf{v}^{\parallel} = a\mathbf{u}$ and $\left\langle \mathbf{v}^{\perp}, \mathbf{v}^{\parallel} \right\rangle = 0$. Note that if we identify a, then the component $\mathbf{v}^{\parallel}$ is known and so is $\mathbf{v}^{\perp}$. Taking inner products with $\mathbf{u}$ on both sides we get

$$\langle \mathbf{v}, \mathbf{u} \rangle = \left\langle \mathbf{v}^{\parallel} + \mathbf{v}^{\perp}, \mathbf{u} \right\rangle = a \langle \mathbf{u}, \mathbf{u} \rangle + \frac{1}{a} \left\langle \mathbf{v}^{\perp}, \mathbf{v}^{\parallel} \right\rangle = a \langle \mathbf{u}, \mathbf{u} \rangle + 0 \Longrightarrow a = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle}.$$

Thus, there exists a unique choice of a for which this decomposition of a generic vector $\mathbf{v}$ holds. The resulting component $\mathbf{v}^{\parallel}$ is known as the orthogonal projection of $\mathbf{v}$ onto $\mathbf{u}$. We include its form in the following definition.

Definition 8.5 — Orthogonal projection (onto a vector). Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space and let $\mathbf{u} \in V \setminus \{\mathbf{0}\}$. The orthogonal projection of $\mathbf{v} \in V$ onto $\mathbf{u}$, denoted by $\mathbf{v}_{\mathbf{u}}^{\parallel}$, is the vector

$$\mathbf{v}_{\mathbf{u}}^{\parallel} := \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle}\mathbf{u}.$$

Note that we changed the notation from $\mathbf{v}^{\parallel}$ to $\mathbf{v}_{\mathbf{u}}^{\parallel}$, as we want to emphasise that the orthogonal projection depends on the vector we project on. Using this notation, the decomposition of $\mathbf{v}$ is

$$\mathbf{v} = \mathbf{v}_{\mathbf{u}}^{\parallel} + \mathbf{v}_{\mathbf{u}}^{\perp},$$

where $\mathbf{v}_{\mathbf{u}}^{\perp} \perp \mathbf{v}_{\mathbf{u}}^{\parallel}$ is given by

$$\mathbf{v}_{\mathbf{u}}^{\perp} = \mathbf{v} - \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle}\mathbf{u}.$$

Note that we can also write the orthogonal projection using the angle α between $\mathbf{v}$ and $\mathbf{u}$:

$$\mathbf{v}_{\mathbf{u}}^{\parallel} := \frac{\|\mathbf{v}\| \cos\alpha}{\|\mathbf{u}\|}\mathbf{u} = \|\mathbf{v}\| \cos\alpha\widehat{\mathbf{u}}.$$

Definition 8.5 is essentially the statement provided for the Euclidean space $\mathbb{E}^3$ in Lecture 1, although we note that it allows also for the projection of the vector $\mathbf{0}$ onto any vector $\mathbf{u}$, which is $\mathbf{0}$. A further comparison to the Euclidean space is included in the next result: in $\mathbb{E}^3$, the segment perpendicular to the direction line of $\mathbf{u}$ had the least length among all segments drawn between $\mathbf{v}$ and the direction line of $\mathbf{u}$. This is the case also in inner product spaces.

Proposition 8.5 Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space and let $\mathbf{u} \in V \setminus \{\mathbf{0}\}$. Let $\mathbf{v} = \mathbf{v}_{\mathbf{u}}^{\parallel} + \mathbf{v}_{\mathbf{u}}^{\perp}$, where $\mathbf{v}_{\mathbf{u}}^{\parallel}$ is the projection of $\mathbf{v}$ onto $\mathbf{u}$. Then

$$\left\|\mathbf{v}_{\mathbf{u}}^{\perp}\right\| = \left\|\mathbf{v} - \mathbf{v}_{\mathbf{u}}^{\parallel}\right\| \le \|\mathbf{v} - \mathbf{z}\| \quad \text{for all } \mathbf{z} \in U := \operatorname{span}\{\mathbf{u}\}.$$

Proof. Let $\mathbf{z} = \mathbf{v}_{\mathbf{u}}^{\parallel} + \mathbf{e}$, for some $\mathbf{e} \in U$. Note that we can write $\mathbf{z}$ in this form since $\mathbf{v}_{\mathbf{u}}^{\parallel} \in U$ also. With this notation, we note that

$$\mathbf{v} - \mathbf{z} = \mathbf{v} - \mathbf{v}_{\mathbf{u}}^{\parallel} - \mathbf{e} = \mathbf{v}_{\mathbf{u}}^{\perp} - \mathbf{e},$$

where $\mathbf{v}_{\mathbf{u}}^{\perp} \perp \mathbf{e}$, since $\mathbf{e} \in U$ is a multiple of $\mathbf{u}$. Therefore, we can apply the Pythagoras theorem:

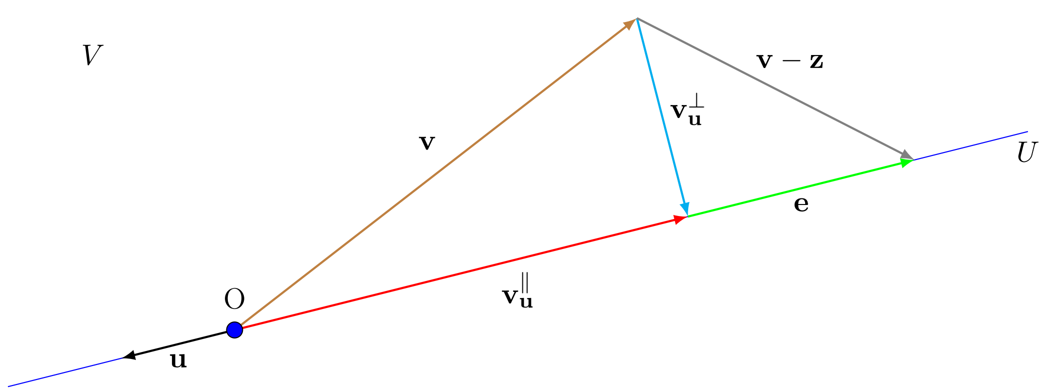
$$\left\|\mathbf{v}_{\mathbf{u}}^{\perp} - \mathbf{e}\right\|^2 = \left\|\mathbf{v}_{\mathbf{u}}^{\perp}\right\|^2 + \|\mathbf{e}\|^2 .$$

Hence,

$$\|\mathbf{v} - \mathbf{z}\|^2 = \left\|\mathbf{v}_{\mathbf{u}}^{\perp} - \mathbf{e}\right\|^2 = \left\|\mathbf{v}_{\mathbf{u}}^{\perp}\right\|^2 + \|\mathbf{e}\|^2 \ge \left\|\mathbf{v}_{\mathbf{u}}^{\perp}\right\|^2 = \left\|\mathbf{v} - \mathbf{v}_{\mathbf{u}}^{\parallel}\right\|^2,$$

with equality holding if and only if $\mathbf{e} = \mathbf{0}$, i.e., if and only if $\mathbf{z} = \mathbf{v}_{\mathbf{u}}^{\parallel}$. ■

R The figure below is included for illustration: the subspace U is a line in the plane V. The proposition simply states that the length $\left\|\mathbf{v}_{\mathbf{u}}^{\perp}\right\|$ of the perpendicular segment is shorter than the length $\|\mathbf{v} - \mathbf{z}\|$ of any other segment drawn from the tip of $\mathbf{v}$ to the line U. Note also that the proof highlights at the end the uniqueness of the segment of least length.



Orthogonal projection $\mathbf{v}_{\mathbf{u}}$ of generic $\mathbf{v}$ in (the plane) V on $\mathbf{u} \in U$.

In the previous result, the vector $\mathbf{v}_{\mathbf{u}}^{\perp}$ is orthogonal to $\mathbf{u}$ and therefore to any multiple of $\mathbf{u}$, i.e., $\mathbf{v}_{\mathbf{u}}^{\perp}$ is orthogonal to any element in the subspace $U = \operatorname{span}\{\mathbf{u}\}$ of V. This suggests a more general definition of orthogonality.

Definition 8.6 — Orthogonal projection (onto a subspace). Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space and let U be a subspace of V. We say $\mathbf{v}$ is orthogonal to U if $\mathbf{v} \perp \mathbf{u}$ for all $\mathbf{u} \in U$. We write $\mathbf{v} \perp U$.

Proposition 8.6 Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space and let U be a subspace of V. Let S be a spanning set for U. Then
$$\mathbf{v} \perp U \iff \mathbf{v} \perp S.$$

Proof. Exercise. ∎

The choice of spanning set in the previous proposition can be a convenient one, for example, an orthogonal basis. Indeed, we will make this choice in the proof of the following result.

We are now ready to extend the concept of orthogonal projection introduced in Definition 8.5.

Theorem 8.7 Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space and let U be a subspace of V. Let $\mathbf{v} \in V$. Then there exists a unique vector $\mathbf{v}_U^{\parallel} \in U$ such that $\mathbf{v}_U^{\perp} := \mathbf{v} - \mathbf{v}_U^{\parallel} \perp U$.

Proof. Let $B = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ be an orthogonal basis for U. For any $\mathbf{v} \in V$, there exists a vector $\mathbf{v}_U^{\parallel}$ in U, given by
$$\mathbf{v}_U^{\parallel} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_k\mathbf{u}_k, \qquad \text{where } a_i := \frac{\langle \mathbf{v}, \mathbf{u}_i\rangle}{\langle \mathbf{u}_i, \mathbf{u}_i\rangle} \quad (i = 1, \dots, k),$$
such that $\mathbf{v} - \mathbf{v}_U^{\parallel} = \mathbf{v}_U^{\perp} \perp U$. To see this, we use Proposition 8.6 and check that $\mathbf{v}_U^{\perp} \perp \mathbf{u}_j$ for all j:
$$\left\langle \mathbf{v}_U^{\perp}, \mathbf{u}_j \right\rangle = \langle \mathbf{v}, \mathbf{u}_j\rangle - \left\langle \mathbf{v}_U^{\parallel}, \mathbf{u}_j \right\rangle = \langle \mathbf{v}, \mathbf{u}_j\rangle - \sum_{i=1}^{k} a_i \langle \mathbf{u}_i, \mathbf{u}_j\rangle = \langle \mathbf{v}, \mathbf{u}_j\rangle - \frac{\langle \mathbf{v}, \mathbf{u}_j\rangle}{\langle \mathbf{u}_j, \mathbf{u}_j\rangle} \langle \mathbf{u}_j, \mathbf{u}_j\rangle = 0.$$

To prove uniqueness, assume there exists $\widetilde{\mathbf{v}}_U^{\parallel} \in U$ satisfying $\mathbf{v} - \widetilde{\mathbf{v}}_U^{\parallel} \perp U$. Let $\widetilde{\mathbf{v}}_U^{\parallel} = \mathbf{v}_U^{\parallel} + \mathbf{e}$ for some $\mathbf{e} \in U$ which we represent in the basis B as
$$\mathbf{e} = \sum_{i=1}^{k} c_i \mathbf{u}_i, \quad c_i = \frac{\langle \mathbf{e}, \mathbf{u}_i\rangle}{\langle \mathbf{u}_i, \mathbf{u}_i\rangle}.$$

Under our assumption, we must have for all j
$$0 = \left\langle \mathbf{v} - \widetilde{\mathbf{v}}_U^{\parallel}, \mathbf{u}_j \right\rangle = \left\langle \mathbf{v} - \mathbf{v}_U^{\parallel} - \mathbf{e}, \mathbf{u}_j \right\rangle = 0 - \langle \mathbf{e}, \mathbf{u}_j\rangle = -c_j \|\mathbf{u}_j\|^2 \implies c_j = 0 \implies \mathbf{e} = \mathbf{0}.$$
∎

8.3 Orthogonal decompositions

The result of the previous proposition confirms that we can write any vector $\mathbf{v}$ in an inner product space as the sum of two orthogonal vectors: $\mathbf{v}_U^{\parallel}$ in U and $\mathbf{v}_U^{\perp}$ in a set disjoint from U. We make this observation more precise via the following definition.

Definition 8.7 — Orthogonal complement. Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space and let U be a subspace of V. The orthogonal complement of U in V is denoted by $U^{\perp}$ and is defined to be the set of vectors in V perpendicular to U
$$U^{\perp} := \{\mathbf{v} \in V : \mathbf{v} \perp U\}.$$

This definition implies that $\mathbf{v}_U^{\perp} \in U^{\perp}$. One can immediately establish the following properties of $U^{\perp}$.

Proposition 8.8 Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space and let U be a subspace of V. Then $U^\perp$ is a subspace of V and $U \cap U^\perp = \{\mathbf{0}\}$.

Proof. The first statement follows from the Subspace criterion 1. The second statement follows by contradiction: if $\mathbf{0} \neq \mathbf{w} \in U \cap U^\perp$, then we must have $\mathbf{w} \perp \mathbf{w}$, i.e., $\langle \mathbf{w}, \mathbf{w}\rangle = 0$, so that $\mathbf{w} = \mathbf{0}$. ■

We end with the following edifying result.

Proposition 8.9 Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space and let U be a subspace of V. Then $V = U \oplus U^\perp$.

Proof. The result follows by the Direct sum criterion 1, since U and $U^\perp$ are two subspaces of V satisfying $V = U + U^\perp$ (why?) and $U \cap U^\perp = \{\mathbf{0}\}$. ■

This is an example of an orthogonal decomposition of a vector space.

LECTURE 9

Gram-Schmidt orthogonalisation

9.1 Orthogonalisation

It is clear from previous lectures that the availability of an orthogonal basis allows for convenient evaluations or derivations of results. However, it is not immediately clear how one can come across such a useful set: often, standard bases fail to be orthogonal with respect to a given inner product.

Example 9.1 Let $V = \mathcal{P}_n([-1,1])$ be the space of polynomials defined on the interval $[-1,1]$ and consider the following inner product:

$$\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)\mathrm{d}x.$$

Then the elements of the power basis $B = \{1, x, \ldots, x^n\}$ satisfy

$$\left\langle x^i, x^j\right\rangle = \int_{-1}^{1} x^{i+j}\mathrm{d}x = \left[\frac{x^{i+j+1}}{i+j+1}\right]_{-1}^{1} = \frac{1-(-1)^{i+j+1}}{i+j+1} = 0, \qquad \text{provided } i+j \text{ is odd.}$$

To devise an orthogonal basis for a finite dimensional inner product space V, we can employ the concept of orthogonal decompositions. To see how this works, Let U be a subspace of an inner product space V spanned by two non-orthogonal vectors $\mathbf{u}, \mathbf{v}$. Consider now the orthogonal decomposition

$$\mathbf{v} = \mathbf{v}_{\mathbf{u}}^{\parallel} + \mathbf{v}_{\mathbf{u}}^{\perp}.$$

Then $\left\{\mathbf{u}, \mathbf{v}_{\mathbf{u}}^{\perp}\right\}$ is an orthogonal set. Moreover, $\operatorname{span}\{\mathbf{u}, \mathbf{v}\} = \operatorname{span}\left\{\mathbf{u}, \mathbf{v}_{\mathbf{u}}^{\perp}\right\}$. As an orthogonal spanning set, $\left\{\mathbf{u}, \mathbf{v}_{\mathbf{u}}^{\perp}\right\}$ is a basis for U. In other words, we generated an orthogonal basis for U, starting from a generic basis. Can we generalise this approach? Consider the subspace spanned by the non-orthogonal vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$. Let $W = \operatorname{span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$. By the previous discussion, we also have $W = \operatorname{span}\left\{\mathbf{u}, \mathbf{v}^{\perp}, \mathbf{w}\right\}$. Moreover, U is a subspace of W so that we can use Theorem 8.7 to write down the orthogonal decomposition

$$\mathbf{w} = \mathbf{w}_U^{\parallel} + \mathbf{w}_U^{\perp},$$

where, by the proof of the theorem,

$$\mathbf{w}_U^{\perp} = \mathbf{w} - \frac{\langle \mathbf{w}, \mathbf{u}\rangle}{\langle \mathbf{u}, \mathbf{u}\rangle} - \frac{\langle \mathbf{w}, \mathbf{v}\ \perp\rangle}{\langle \mathbf{v}^{\perp}, \mathbf{v}^{\perp}\rangle}.$$

As before, we note that span$\{\mathbf{u}, \mathbf{v}^{\perp}, \mathbf{w}^{\perp}\}$ is a spanning orthogonal set, hence a basis for W.

Assume now that U is a proper subspace of V and (somehow) we already have an orthogonal basis B for U. Since the orthogonal complement U^T contains vectors orthogonal to U, these will also be orthogonal to the elements in B. We can then select one such vector and add it to the set B. This will result in an orthogonal basis for a subspace of increased dimension (by one). Continuing in this way, we eventually generate an orthogonal set of dimesion n, which will then have to be a basis for V. This procedure is guaranteed to work due to the existence and uniqueness result of Theorem 8.7. An iterative procedure for the above approach is included below.

Algorithm: generic orthogonalisation: compute orthogonal basis B_n

choose any basis $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of V
define $B_1 = \{\mathbf{v}_1\}$
for $k = 1, 2, \ldots, n-1$
  define $U_k = \text{span} B_k$
  generate $\mathbf{u}_k \in U_k^{\perp}$ using $\mathbf{v}_k$
  define $B_{k+1} = B_k \cup \{\mathbf{u}_k\}$
end
return orthogonal basis B_n for V

Note that the output B_n of the above algorithm is an orthogonal set with n elements, as we append to the initial (single-vector) set B_1 another $n-1$ vectors. Hence B_n is a basis.

R If we run the above algorithm for $k = m < n-1$ steps, we will construct an orthogonal basis B_m for a subspace of U_m of V.

9.2 Gram-Schmidt procedure

The only instruction that needs to be described is that referring to the construction of the vectors $\mathbf{u}_k$. By Theorem 8.7, in order to construct a vector orthogonal to U we need to

- choose a vector $\mathbf{v}$;
- subtract from it the orthogonal projections onto each of the basis elements of U.

In the case of the Gram-Schmidt procedure for orthogonalisation, the vector $\mathbf{v}$ is chosen to be $\mathbf{v}_k$, while the basis elements of U are chosen to be the previously computed orthogonal vectors. The resulting procedure is known as the Gram-Schmidt procedure (or process).

Algorithm: Gram-Schmidt orthogonalisation: compute orthogonal basis B_n

choose any basis $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of V
define $B_1 = \{\mathbf{v}_1\}$
for $k = 1, 2, \ldots, n-1$

$$\mathbf{u}_k = \mathbf{v}_k - \sum_{i=1}^{k-1} \frac{\langle \mathbf{v}_k, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$$

  define $B_{k+1} = B_k \cup \{\mathbf{u}_k\}$
end
return orthogonal basis B_n for V

We can re-write the above algorithm to include normalisation of the new basis vectors. This results in a somewhat simplified algorithm, due to the fact that the previously generated basis elements are unit vectors, so that $\langle \mathbf{u}_i, \mathbf{u}_i \rangle = 1$.

Algorithm: Gram-Schmidt orthonormalisation: compute orthonormal basis B_n

choose any basis $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of V
define $B_1 = \{\mathbf{v}_1\}$
for $k = 1, 2, \ldots, n-1$

$$\mathbf{u}_k = \mathbf{v}_k - \sum_{i=1}^{k-1} \langle \mathbf{v}_k, \widehat{\mathbf{u}}_i \rangle \, \widehat{\mathbf{u}}_i$$

define $\widehat{\mathbf{u}}_k = \mathbf{u}_k / \|\mathbf{u}_k\|$
define $B_{k+1} = B_k \cup \{\widehat{\mathbf{u}}_k\}$
end
return orthonormal basis B_n for V

Let us write the algorithm explicitly, indicating the first few steps and also the generic step from the above algorithm.

$$\begin{cases}
\mathbf{u}_1 = \mathbf{v}_1 & \widehat{\mathbf{u}}_1 := \mathbf{u}_1 / \|\mathbf{u}_1\|, \\
\mathbf{u}_2 = \mathbf{v}_2 - \langle \mathbf{v}_2, \widehat{\mathbf{u}}_1 \rangle \, \widehat{\mathbf{u}}_1 & \widehat{\mathbf{u}}_2 := \mathbf{u}_2 / \|\mathbf{u}_2\|, \\
\mathbf{u}_3 = \mathbf{v}_3 - \langle \mathbf{v}_3, \widehat{\mathbf{u}}_1 \rangle \, \widehat{\mathbf{u}}_1 - \langle \mathbf{v}_3, \widehat{\mathbf{u}}_2 \rangle \, \widehat{\mathbf{u}}_2 & \widehat{\mathbf{u}}_3 := \mathbf{u}_3 / \|\mathbf{u}_2\|, \\
\qquad \cdots\cdots\cdots\cdots\cdots\cdots & \\
\mathbf{u}_k = \mathbf{v}_k - \displaystyle\sum_{i=1}^{k-1} \langle \mathbf{v}_k, \widehat{\mathbf{u}}_i \rangle \, \widehat{\mathbf{u}}_i & \widehat{\mathbf{u}}_k := \mathbf{u}_k / \|\mathbf{u}_k\|, \\
\qquad \cdots\cdots\cdots\cdots\cdots\cdots & \\
\mathbf{u}_n = \mathbf{v}_n - \displaystyle\sum_{i=1}^{n-1} \langle \mathbf{v}_n, \widehat{\mathbf{u}}_i \rangle \, \widehat{\mathbf{u}}_i & \widehat{\mathbf{u}}_n := \mathbf{u}_n / \|\mathbf{u}_n\|.
\end{cases}$$

We note that at each step we need to use quantities (unit basis elements) computed in previous steps. Finally, we note that

$$\mathbf{v}_k = \mathbf{u}_k + \sum_{i=1}^{k-1} \langle \mathbf{v}_k, \widehat{\mathbf{u}}_i \rangle \, \widehat{\mathbf{u}}_i = \|\mathbf{u}_k\| \, \widehat{\mathbf{u}}_k + \sum_{i=1}^{k-1} \langle \mathbf{v}_k, \widehat{\mathbf{u}}_i \rangle \, \widehat{\mathbf{u}}_i = \sum_{i=1}^{k} \langle \mathbf{v}_k, \widehat{\mathbf{u}}_i \rangle \, \widehat{\mathbf{u}}_i,$$

since $\|\mathbf{u}_k\| = \langle \mathbf{u}_k, \widehat{\mathbf{u}}_k \rangle = \langle \mathbf{v}_k, \widehat{\mathbf{u}}_k \rangle$. Hence, every $\mathbf{v}_k$ can be written as a linear combination of only the first k vectors in the newly computed orthonormal basis. We will discuss the significance of this later, when we consider orthogonalisation in $\mathbb{R}^n$.

We end this lecture with two examples involving standard inner product spaces.

9.3 Examples

Example 9.2 Consider the following non-orthogonal basis for $\mathbb{R}^3$

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

Let $\langle\cdot,\cdot\rangle$ denote the Euclidean inner-product. Using the Gram-Schmidt iterative procedure we find

$$\mathbf{u}_1 = \mathbf{v}_1, \quad \widehat{\mathbf{u}}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\\0\end{bmatrix};$$

$$\mathbf{u}_2 = \mathbf{v}_2 - \langle\mathbf{v}_2, \widehat{\mathbf{u}}_1\rangle\,\widehat{\mathbf{u}}_1 = \begin{bmatrix}1\\1\\1\end{bmatrix} - \begin{bmatrix}1\\1\\0\end{bmatrix} = \begin{bmatrix}0\\0\\1\end{bmatrix}, \quad \widehat{\mathbf{u}}_2 = \begin{bmatrix}0\\0\\1\end{bmatrix};$$

$$\mathbf{u}_3 = \mathbf{v}_3 - \langle\mathbf{v}_3, \widehat{\mathbf{u}}_1\rangle\,\widehat{\mathbf{u}}_1 - \langle\mathbf{v}_3, \widehat{\mathbf{u}}_2\rangle\,\widehat{\mathbf{u}}_2 = \begin{bmatrix}0\\1\\1\end{bmatrix} - \frac{1}{2}\begin{bmatrix}1\\1\\0\end{bmatrix} - \begin{bmatrix}0\\0\\1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}-1\\1\\0\end{bmatrix}, \quad \widehat{\mathbf{u}}_3 = \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\1\\0\end{bmatrix}.$$

The above example confirms that orthonormal bases are not unique: the canonical basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is also orthonormal with respect to the Euclidean inner product.

Example 9.3 Consider the power basis $\{1, x, x^2\} =: \{p_1, p_2, p_3\}$ for $\mathcal{P}_2([-1,1])$. Let us apply the Gram-Schmidt procedure without normalisation using the inner product

$$\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)\mathrm{d}x.$$

We will denote the new basis by q_1, q_2, q_3. We find

$$q_1 = p_1 = 1;$$

$$q_2 = p_2 - \frac{\langle p_2, q_1\rangle}{\langle q_1, q_1\rangle}q_1 = x$$

since

$$\langle p_2, q_1\rangle = \int_{-1}^{1} x\mathrm{d}x = 0$$

and finally

$$q_3 = p_3 - \frac{\langle p_3, q_2\rangle}{\langle q_2, q_2\rangle}q_2 - \frac{\langle p_3, q_1\rangle}{\langle q_1, q_1\rangle}q_1 = x^2 - \frac{1}{3}$$

since

$$\langle p_3, q_2\rangle = \int_{-1}^{1} x^3\mathrm{d}x = 0, \quad \langle p_3, q_1\rangle = \int_{-1}^{1} x^2 = \frac{2}{3}, \quad \langle q_1, q_1\rangle = \int_{-1}^{1} 1\mathrm{d}x = 2.$$

Therefore, the resulting orthogonal basis is $\{1, x, x^2 - \frac{1}{3}\}$.

LECTURE 10

Spotlight: $\mathbb{R}^n$

The vector space $\mathbb{R}^n$ provides a natural setting for many problems of interest. In this lecture, we will take a detailed view at $\mathbb{R}^n$, as a vector space over $\mathbb{R}$. Some additional objects, concepts and definitions will be needed, as we aim to provide a complete picture in light of the material studied so far.

10.1 Bakground and notation

10.1.1 Vectors

Vectors in $\mathbb{R}^n$ are denoted generically by $\mathbf{v}$ and are defined to be **column vectors** with n real entries:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

The ith entry of $\mathbf{v}$ is denoted by $[\mathbf{v}]_i$ or v_i; we also indicate $\mathbf{v}$ using a generic entry in square brackets: $\mathbf{v} = [v_i]_{1\le i\le n}$, or simply $\mathbf{v} = [\mathbf{v}_i]$ if the range for i is clear from the context; this is useful also when the entries are the output of a function of i, e.g.,

$$\mathbf{v} = [i^2 - 1]_{1\le i\le 3} = \begin{bmatrix} 0 \\ 3 \\ 8 \end{bmatrix}.$$

We will use the convention that $\mathbf{0}$ is the vector of zeros and $\mathbf{1}$ is the vector of ones. We will also come across **row vectors**, which will be viewed as the transpose of some column vector, with the entries written in a horizontal list; the notation is given by superscript T:

$$\mathbf{r} = \mathbf{v}^T = [v_1, v_2, \ldots, v_n]$$

We will ocasionally need to indicate that a row vector belongs to a different set denoted by $R^{1\times n}$. The reason for this notation is that both column vectors $\mathbf{v} \in \mathbb{R}^n =: \mathbb{R}^{n\times 1}$ and row vectors $\mathbf{r} \in \mathbb{R}^{1\times n}$ are special cases of **matrices** which are elements of the set of rectangular arrays of size $n \times m$, denoted by $\mathbb{R}^{n\times m}$:

- a column vector is a matrix with n rows and 1 column;
- a row vector is a matrix with 1 row and n columns.

10.1.2 Matrices

At this stage, we will find it useful to introduce and work with matrices: a first reason is that one can view matrices as storing a set of column vectors from $\mathbb{R}^n$. Matrices, denoted by A, are viewed as elements of the set of retangular $n \times m$ arrays of real numbers with n rows and m columns denoted by $\mathbb{R}^{n\times m}$:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2m} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{im} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nj} & \cdots & a_{nm} \end{bmatrix}.$$

We write

$$[A]_{ij} = [A]_{i,j} = a_{ij} = a_{i,j}, \quad A = [a_{ij}] = [a_{i,j}],$$

where $1 \le i \le n, 1 \le j \le m$.
We will often work with or refer to the rows and columns of a matrix; their definitions are included below.

Definition 10.1 Let $1 \le j \le n$. **The jth column of a matrix** $A \in \mathbb{R}^{n\times m}$ is the column vector $\mathbf{c}_j(A) \in \mathbb{R}^n$ defined below

$$[\mathbf{c}_j(A)]_i := [a_{ij}]_{1\le i\le n}, \quad \mathbf{c}_j(A) = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix}.$$

Definition 10.2 Let $1 \le i \le m$. **The ith row of a matrix** $A \in \mathbb{R}^{n\times m}$ is the row vector $\mathbf{r}_i(A) \in \mathbb{R}^{1\times m}$ defined below

$$[\mathbf{r}_i(A)]_j := [a_{ij}]_{1\le j\le m}, \quad \mathbf{r}_i(A) = \begin{bmatrix} a_{i1} \; a_{i2} \; \ldots \; a_{im} \end{bmatrix}.$$

With the above definitions in place, a matrix $A \in \mathbb{R}^{m\times n}$ can be given a row or column representation:

$$A = \begin{bmatrix} \mathbf{r}_1(A) \\ \mathbf{r}_2(A) \\ \vdots \\ \mathbf{r}_n(A) \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1(A) & \mathbf{c}_2(A) & \cdots & \mathbf{c}_m(A) \end{bmatrix}.$$

Special matrices

We list below a small subset of special matrices.

Zero matrix: Denoted by $O_{m,n}$ or simply by O, this is the matrix with all entries equal to zero.
Identity matrix: This is usually assumed to be a square matrix denoted by $I = I_n = [\delta_{ij}] \in \mathbb{R}^{n\times n}$, where δ_{ij} is the Kronecker delta symbol:

$$\delta_{ij} = \begin{cases} 0 & i \ne j, \\ 1 & i = j. \end{cases}$$

The ith column of I_n is the i**th canonical vector** $\mathbf{e}_i := \mathbf{c}_i(I_n) \in \mathbb{R}^n$.
Diagonal matrices have the general form $A = [a_{ii}\delta_{ij}]$. Note that for any diagonal matrix $\mathbf{c}_j(A) = a_{jj}\mathbf{e}_j$.
Triangular matrices: A matrix A is **lower triangular** if $a_{ij} = 0$ for $i < j$. A matrix A is said to be

an **upper triangular matrix** if $a_{ij} = 0$ for $i > j$. For example, for $m = n$, a general lower triangular matrix has the form

$$\begin{bmatrix} a_{11} & & & & \\ a_{21} & a_{22} & & & \\ a_{31} & a_{32} & a_{33} & & \\ \vdots & \vdots & \ddots & \ddots & \\ a_{m1} & a_{m2} & \cdots & a_{m,m-1} & a_{mm} \end{bmatrix}.$$

Matrix operations

We define below various operations with matrices. For simplicity, we give the definitions for the case of real matrices. We start with the operations that make $\mathbb{R}^{n\times m}$ into a vector space.

Vector space operations

- **Multiplication by scalars** Let $\alpha \in \mathbb{R}, A \in \mathbb{R}^{n\times m}$. This operation is defined via $[\alpha A]_{ij} := \alpha a_{ij}$. It satisfies the corresponding vector space requirements.
- **Matrix addition** Let $A, B \in \mathbb{R}^{n\times m}$. This operation is defined via $[A + B]_{ij} := a_{ij} + b_{ij}$. This operation is associative, commutative and it has an additive identity (the zero matrix) and an additive inverse for each matrix A (the matrix $-A$).

Functional operations

Most of the operations defined below will arise naturally when we study linear transformations in Part II. At this stage, it is preferrable to introduce notation and a few definitions, some of which are already familiar.

- **Transposition** Let $A \in \mathbb{R}^{m\times n}$. We define the transposition operator $t : \mathbb{R}^{m\times n} \to \mathbb{R}^{n\times m}$ as the matrix $t(A) \in \mathbb{R}^{n\times m}$ with entries $[t(A)]_{ij} = a_{ji}$. This is also known as the transpose of A and is denoted by A^T. The following observations hold:
  - Transposition of a scalar (viewed as a 1-by-1 matrix) is the scalar itself.
  - Transposition of a column vector is a row vector, e.g., $\mathbf{v}^T = [v_1\ v_2\ \cdots\ v_n]$.
  - Transposition of a transposed matrix is the matrix itself: $(A^T)^T = A$.
  - Transposition of a matrix-matrix product: $(AB)^T = B^T A^T$.
- **Matrix-matrix multiplication** Let $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times q}$. The product AB is defined to be a matrix $C \in \mathbb{R}^{m\times q}$ with entries

  $$[C]_{ij} := \sum_{k=1}^{n} a_{ik} b_{kj}.$$

  The following special cases give rise to certain standard operations.
  - $m = q = 1$: inner product (of column vectors);
  - $q = 1$: matrix-vector product;
  - $m = 1$: vector-matrix product;
  - $n = 1$: outer product (of column vectors);

  We provide descriptions below for all these operations.
- **Inner-product**: If $m = q = 1$, the matrices $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times q}$ become row and column vectors, respectively, while the matrix C defaults to a scalar. We represent the row vector as the transpose of a column vector, say $A = \mathbf{v}^T$, with $B = \mathbf{w}$. Given $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, the **inner-product** of $\mathbf{v}, \mathbf{w}$ is denoted by $\langle \mathbf{v}, \mathbf{w}\rangle$ is defined as the matrix-matrix product of $\mathbf{v}^T$ and $\mathbf{w}$:

  $$\langle \mathbf{v}, \mathbf{w}\rangle := \mathbf{v}^T\mathbf{w} = \sum_{i=1}^{n} v_i w_i \in \mathbb{R}.$$

  This is indeed the Euclidean inner product that we encountered in earlier lectures.

- **Matrix-vector product** If $q = 1$, the matrix $B \in \mathbb{R}^{n\times q}$ becomes a column vector, say $B = \mathbf{v}$ and the matrix-matrix product yields a column vector whose entries are inner-products; it can also be seen to be a linear combination of the columns of A:

$$[A\mathbf{v}]_i := \sum_{j=1}^{n} a_{ij}v_j = \begin{cases} \sum_{j=1}^{n}[\mathbf{r}_i(A)]_j v_j = \mathbf{r}_i(A)\mathbf{v}, \\ \sum_{j=1}^{n}[\mathbf{c}_j(A)]_i v_j = \left[\sum_{j=1}^{n} v_j\mathbf{c}_j(A)\right]_i \end{cases}$$

so that

$$A\mathbf{v} = \begin{bmatrix} \mathbf{r}_1(A)\mathbf{v} \\ \vdots \\ \mathbf{r}_m(A)\mathbf{v} \end{bmatrix}, \qquad A\mathbf{v} = \sum_{j=1}^{n} v_j\mathbf{c}_j(A).$$

R The last expression above indicates that the matrix-vector product is a linear combination of the columns of A. We will use this fact later.

- **Vector-matrix product** When $m = 1$, the matrix $A \in \mathbb{R}^{m\times n}$ becomes a row vector, say $A = \mathbf{w}^T$ and the matrix-matrix product yields a row vector whose entries are scalar products; it can also be seen to be a linear combination of the rows of B:

$$[\mathbf{w}^T B]_j := \sum_{i=1}^{m} w_i b_{ij} = \begin{cases} \sum_{i=1}^{m} w_i[\mathbf{c}_j(B)]_i = \mathbf{w}^T\mathbf{c}_j(B) \\ \sum_{i=1}^{m} w_i[\mathbf{r}_i(B)]_j = \left[\sum_{i=1}^{m} w_i\mathbf{r}_i(B)\right]_j \end{cases}$$

so that

$$\mathbf{w}^T B = [\mathbf{w}^T\mathbf{c}_1(B), \cdots, \mathbf{w}^T\mathbf{c}_n(B)], \quad \mathbf{w}^T B = \sum_{i=1}^{m} w_i\mathbf{r}_i(B).$$

- **Outer product** When $n = 1$, the matrices $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times q}$ become column and row vectors, respectively, while the matrix C has size m-by-q. As before, re-labeling $A = \mathbf{v}, B = \mathbf{w}^T$, with $\mathbf{v} \in \mathbb{R}^m, \mathbf{w} \in \mathbb{R}^q$; we define the outer product as the $m \times q$ matrix denoted by $\mathbf{v}\mathbf{w}^T \in \mathbb{R}^{m\times q}$

$$\mathbf{v}\mathbf{w}^T := \begin{bmatrix} v_1w_1 & v_1w_2 & \cdots & v_1w_q \\ v_2w_1 & v_2w_2 & \cdots & v_2w_q \\ \vdots & \vdots & \vdots & \vdots \\ v_mw_1 & v_mw_2 & \cdots & v_mw_q \end{bmatrix}.$$

In other words, $[\mathbf{v}\mathbf{w}^T]_{ij} = v_iw_j$ for $1 \le i \le m, 1 \le j \le q$.
- **Matrix inverse** Let A be a square matrix. We say A has an inverse if there exists a square matrix B such that $AB = BA = I$. The inverse is denoted by A^{-1}. If the inverse exists, it is unique; we say A **is invertible or non-singular**. Otherwise, we say A **is not invertible or singular**.

R When A is rectangular, we can only define a one-sided inverse (or pseudo-inverse): this is only possible if A has full rank (we discuss rank later).

10.2 Subspaces. Linear independence. Bases.

Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ denote a set of vectors in $\mathbb{R}^n$. Let $A = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_k] \in \mathbb{R}^{n\times k}$. Recall that a linear combination of $\mathbf{v}_i$ can be written as the matrix-vector product:

$$A\mathbf{c} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k,$$

where $\mathbf{c} \in \mathbb{R}^k$ is the vector of coefficients. This observation justifies the following definition.

Definition 10.3 — Column space of a matrix. Let $A \in \mathbb{R}^{n\times k}$. The column space of A, denoted by col A is the subspace of $\mathbb{R}^n$ spanned by the columns of A:

$$\text{col } A = \text{span}\{\mathbf{c}_1(A), \mathbf{c}_2(A), \ldots, \mathbf{c}_k(A)\}.$$

Since $A\mathbf{c}$ is a linear combination, we can immediately derive the following result based on the definition of linear dependence.

Proposition 10.1 Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ denote a set of vectors in $\mathbb{R}^n$. Let $A = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_k] \in \mathbb{R}^{n\times k}$. The set S is linearly dependent if and only if $A\mathbf{c} = \mathbf{0}$ for all $\mathbf{c} \in \mathbb{R}^n$.

Linear dependence of a set of columns is linked to the concept of rank.

Definition 10.4 — Rank of a matrix. Let $A \in \mathbb{R}^{n\times k}$. The rank of A is the number of linearly independent columns of A.

Using our results on dimension, we immediately derive this result.

Proposition 10.2 Let $A \in \mathbb{R}^{n\times k}$. Then rank $A = \dim$ col A.

It is evident that for any non-zero matrix in $\mathbb{R}^{n\times k}$ we have $1 \leq \text{rank } A \leq k$. This allows for the following terminology.

Definition 10.5 We say $A \in \mathbb{R}^{n\times k}$ has **full rank** if rank $A = k$. Otherwise, A is said to be **rank-deficient**.

We end this section with the matrix interpretation of the concepts of sum and direct sum. Let $B \in \mathbb{R}^{n\times m}$ be written in block form:

$$B = [A_1\ A_2], \quad A_i \in \mathbb{R}^{n\times m_i}, \quad i = 1, 2, \quad \text{with } m_1 + m_2 = m.$$

Then

$$\mathbf{u} = B\mathbf{c} = A_1\mathbf{c}_1 + A_2\mathbf{c}_2 = \mathbf{u}_1 + \mathbf{u}_2.$$

Hence, $V = U_1 + U_2$, where U_i are subspaces spanned by the columns of A_i for $i = 1, 2$. This is a direct sum provided the decomposition is unique. In particular, this holds if col $A_1 \cap$ col $A_2 = \{\mathbf{0}\}$.

A maximal linearly independent set for $\mathbb{R}^n$ is given by the n canonical vectors; these are vectors with only one entry (equal to 1) in the ith position:

$$\mathbf{e}_i = \delta_{ij} = \begin{cases} 0 & i \neq j, \\ 1 & i = j, \end{cases} \qquad i = 1, 2, \ldots, n.$$

Therefore, $\mathbb{R}^n$ is an n-dimensional vector space. We can associate with any basis of $\mathbb{R}^n$ a **square matrix** whose columns are the basis elements, say, $\mathbf{v}_i$:

$$B = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n].$$

In the case of the canonical basis, we obtain $B = I = I_n \in \mathbb{R}^{n\times n}$, the identity matrix of size n. Note that linear independence holds if and only if

$$\mathbf{0} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots c_n\mathbf{v}_n = B\mathbf{c} \Longrightarrow \mathbf{c} = 0.$$

We will see later that if $B\mathbf{c} = \mathbf{0}$ for some non-zero $\mathbf{c}$, then the matrix B is singular, i.e., $\det B = 0$. Hence, the set is linearly independent if and only if B is non-singular.

Example 10.1 Consider the following set of vectors in $\mathbb{R}^3$

$$S = \left\{ \mathbf{v}_1 = \begin{bmatrix}1\\1\\1\end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix}1\\1\\0\end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix}1\\0\\0\end{bmatrix} \right\}$$

Then the matrix $B = [\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3]$ satisfies

$$\det B = \det \begin{bmatrix}1 & 1 & 1\\1 & 1 & 0\\1 & 0 & 0\end{bmatrix} = 1$$

so that S is linearly independent.

10.3 Inner products. Norms.

The set $\mathbb{R}^n$ is an inner product space when equipped with the standard Euclidean inner product

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T \mathbf{w} = \sum_{i=1}^{n} v_i w_i \in \mathbb{R}.$$

The corresponding induced norm is the Euclidean norm

$$\|\mathbf{v}\| = \left(\sum_{i=1}^{n} v_i^2 \right)^{1/2}.$$

This is also known as the **2-norm** and is written as $\|\mathbf{v}\|_2$, as it corresponds to the special case $p = 2$ of the family of norms

$$\|\mathbf{v}\|_p = \left(\sum_{i=1}^{n} |v|_i^p \right)^{1/p}, \qquad p \in \mathbb{N}.$$

Another p-norm corresponds to 'setting $p = \infty$':

$$\|\mathbf{v}\|_\infty = \max_{1 \le i \le n} |v_i| .$$

Note that these norms are not induced by an inner-product; instead, they satisfiy the norm properties listed in the following definition.

Definition 10.6 A norm on a vector space is a mapping $\|\cdot\| : V \to R_+ \cup \{0\}$ satisfying the following properties:

i. $\|\mathbf{u} + \mathbf{v}\| \le \|\mathbf{u}\| + \|\mathbf{v}\|$;
ii. $\|a\mathbf{v}\| = |a|\,\|\mathbf{v}\|$;
iii. $\|\mathbf{v}\| \ge 0$ with $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

Exercise 10.1 Check the norm properties for $\|\cdot\|_1$ and $\|\cdot\|_\infty$.

Let now $B \in \mathbb{R}^{n \times n}$ have full rank. Then its columns form a basis for $\mathbb{R}^n$ and we can write any vector as a linear combination of these columns. This observation leads to an alternative form for the inner product using the coordinates in the chosen basis. Let $\mathbf{v} = B\mathbf{c}, \mathbf{w} = B\mathbf{d}$. Then

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T \mathbf{w} = (B\mathbf{c})^T B\mathbf{d} = \mathbf{c}^T B^T B\mathbf{d} = \mathbf{c}^T H\mathbf{d},$$

where $H = B^T B$ and $\langle \cdot, \cdot \rangle_H$ is the $H-$inner product induced by the matrix H. Correspondingly, we have an $H-$induced norm:

$$\|\mathbf{c}\|_H = \sqrt{\mathbf{c}^T H \mathbf{c}} = \|B\mathbf{c}\|_2 .$$

The concept of orthonormality leads to the concept of orthogonal matrices. If our basis for $\mathbb{R}^n$ is orthonormal, B has orthonormal columns and therefore $B^T B = I_n$. Moreover, $BB^T = I_n$ (why?).

Definition 10.7 A matrix $Q \in \mathbb{R}^{n\times n}$ is said to be orthogonal if $Q^T Q = QQ^T = I_n$.

10.3.1 Gram-Schmidt and the QR decomposition

Let $\{\mathbf{v}_1, \mathbf{v}_2 \dots, \mathbf{v}_n\}$ be a basis for $\mathbb{R}^n$ and let $\{\widehat{\mathbf{u}_1}, \widehat{\mathbf{u}_2}, \dots, \widehat{\mathbf{u}}_n\}$ denote the orthonormal basis obtained by the Gram-Schmidt procedure. Define the matrices

$$A = [\mathbf{v}_1\ \mathbf{v}_2\ \dots\ \mathbf{v}_n], \quad Q = [\widehat{\mathbf{u}_1}\ \ \widehat{\mathbf{u}_2}\ \ \dots\ \ \widehat{\mathbf{u}}_n].$$

Recall now the expression relating $\mathbf{v}_k$ to $\widehat{\mathbf{u}}_i$:

$$\mathbf{v}_k = \mathbf{u}_k + \sum_{i=1}^{k-1} \langle \mathbf{v}_k, \widehat{\mathbf{u}}_i \rangle\, \widehat{\mathbf{u}}_i = \|\mathbf{u}_k\|\, \widehat{\mathbf{u}}_k + \sum_{i=1}^{k-1} \langle \mathbf{v}_k, \widehat{\mathbf{u}}_i \rangle\, \widehat{\mathbf{u}}_i = \sum_{i=1}^{k} \langle \mathbf{v}_k, \widehat{\mathbf{u}}_i \rangle\, \widehat{\mathbf{u}}_i,$$

Let us define $r_{ik} := \langle \mathbf{v}_k, \widehat{\mathbf{u}}_i \rangle$. Then we can identify the following linear combination and write it as a matrix-vector product:

$$\mathbf{v}_k = \sum_{i=1}^{k} r_{ik}\widehat{\mathbf{u}}_i = [\widehat{\mathbf{u}_1}\ \ \widehat{\mathbf{u}_2}\ \ \dots\ \ \widehat{\mathbf{u}}_k] \begin{bmatrix} r_{1k} \\ r_{2k} \\ \vdots \\ r_{kk} \end{bmatrix} = [\widehat{\mathbf{u}_1}\ \ \widehat{\mathbf{u}_2}\ \ \dots\ \ \widehat{\mathbf{u}}_n] \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = Q \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Putting together all the vectors $\mathbf{v}_k$ in the matrix A we obtain the following matrix equality:

$$A = [\mathbf{v}_1\ \mathbf{v}_2\ \dots\ \mathbf{v}_n] = Q \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1k} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2k} & \cdots & r_{2n} \\ 0 & 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & r_{kk} & \ddots & r_{kn} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & r_{nn} \end{bmatrix} = QR,$$

where Q is an orthogonal matrix and R is an upper triangular matrix with positive diagonal entries

$$\mathbf{r}_{kk} = \langle \mathbf{v}_k, \widehat{\mathbf{u}}_k \rangle = \|\mathbf{u}_k\| > 0.$$

Thus, the Gram-Schmidt procedure in $\mathbb{R}^n$ amounts to factorising a generic matrix (possibly rectangular): this known as the **QR factorisation of matrix** A.

II. Linear transformations

LECTURE 11

Overview: linear transformations

The material studied so far, namely vector spaces, is motivated by the need to understand the behaviour of (and work with) functions defined on vector spaces. We will restrict our attention to functions that satisfy a linearity property, also known as linear transformations.

11.1 Definitions, notation, properties.

Definition 11.1 — Linear transformation. Let $V(\mathbb{F}), W(\mathbb{F})$ be vector spaces. A linear transformation is a map $f : V \to W$ satisfying the linearity property

$$f(a\mathbf{u} + b\mathbf{v}) = af(\mathbf{u}) + bf(\mathbf{v}),$$

for all vectors $\mathbf{u}, \mathbf{v} \in V$ and all scalars $a, b \in \mathbb{F}$.

The set of linear transformations is denoted by $\mathcal{L}(V, W)$. We write $f \in \mathcal{L}(V, W)$ to indicate a transformation $f\colon V \to W$ which is linear.

Note that the terms 'transformation' and 'map' will be used interchangeably throughout this course.

R It is important to note that the vector space operations are specific to each of the vector spaces V, W. Using the notation from Lecture 1, given $V(\mathbb{F}) = (V, \color{red}{+}, \color{red}{\bullet}, \mathbb{F})$ and $W(\mathbb{F}) = (W, \color{cyan}{+}, \color{cyan}{\bullet}, \mathbb{F})$, the linearity property reads

$$f(a\color{red}{\bullet}\mathbf{u}\color{red}{+}b\color{red}{\bullet}\mathbf{v}) = a\color{cyan}{\bullet}f(\mathbf{u})\color{cyan}{+}b\color{cyan}{\bullet}f(\mathbf{v}).$$

This is clearly awkward to use: we will indeed continue using the vector space operations in the same way, assuming that distinctions are clear from the context.

We will encounter and study (some of) the following types of linear maps:
- **homomorphisms**: these are just general linear transformations $f : V \to W$;
- **isomorphisms**: these are homomorphisms $f : V \to W$ that are invertible;
- **endomorphisms**: these are linear transformations $f : V \to V$;
- **automorphisms**: these are endomorphisms $f : V \to V$ that are invertible.

In this lecture, we consider the general case of homomorphisms, with later lectures dedicated to the study of the other three types of maps.

Using Definition 11.1, we immediately derive the following properties of linear maps:

$$f(\mathbf{u} \pm \mathbf{v}) = f(\mathbf{u}) \pm f(\mathbf{v}), \qquad f(a\mathbf{v}) = af(\mathbf{v}),$$

and by setting $a = 0$ and $a = -1$, respectively, we obtain

$$f(\mathbf{0}_\mathrm{V}) = \mathbf{0}_\mathrm{W}, \quad f(-\mathbf{v}) = -f(\mathbf{v}),$$

where $\mathbf{0}_\mathrm{V}, \mathbf{0}_\mathrm{W}$ are the zero vectors in V and W, respectively. Note also that the second relation is rigourously written as $f(\mathbf{v}^-) = f(\mathbf{v})^-$.

We will use the following terminology and notation in relation to linear maps $f : V \to W$:

- V is the **domain** of f;
- W is the **codomain** of f;
- $\mathbf{w} = f(\mathbf{v})$ is the **image** of $\mathbf{v} \in V$;
- $\mathbf{v}$ is a **pre-image** of $\mathbf{w} \in W$, if $\mathbf{w} = f(\mathbf{v})$;
- $\mathrm{im}\ f := f(V) := \{\mathbf{w} \in W : \mathbf{w} = f(\mathbf{v}), \mathbf{v} \in V\} \subseteq W$ is the **image** or **range** of f;
- $\mathrm{ker} f = \{\mathbf{v} \in V : f(\mathbf{v}) = \mathbf{0}_\mathrm{W}\} \subseteq V$ is the **kernel** or **nullspace** of f.

We also recall the following properties of general maps.

Definition 11.2 — Injective map. A map $f : V \to W$ is said to be injective (or one-to-one) if for any $\mathbf{v}_1, \mathbf{v}_2 \in V$, $f(\mathbf{v}_1) = f(\mathbf{v}_2)$ implies $\mathbf{v}_1 = \mathbf{v}_2$.

Definition 11.3 — Surjective map. A map $f : V \to W$ is said to be surjective (or onto) if for all $\mathbf{w} \in W$ there exists at least one $\mathbf{v} \in V$ such that $f(\mathbf{v}) = \mathbf{w}$.

Definition 11.4 — Bijective map. A map $f : V \to W$ is said to be bijective if it is injective and surjective.

Definition 11.5 — Trivial kernel. The kernel of a linear map $f : V \to W$ is said to be trivial if $\mathrm{ker} f = \{\mathbf{0}\}$.

In addition to the notation $f(V)$, we will also employ the notation $f(S)$ to indicate the image of a finite set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$:

$$f(S) := \{f(\mathbf{v}_1), f(\mathbf{v}_2), \dots, f(\mathbf{v}_k)\}\,.$$

The following results provide additional descriptions of linear maps.

Proposition 11.1 Let $V(\mathbb{F}), W(\mathbb{F})$ be vector spaces. The set of linear maps $\mathcal{L}(V, W)$ is a vector space over $\mathbb{F}$.

Proof. Exercise. ■

Proposition 11.2 Let $f : V \to W$ be a linear map. Then

i. f is injective if and only if $\mathrm{ker} f = \{\mathbf{0}_\mathrm{V}\}$.

ii. f is surjective if and only if $\mathrm{im}\ f = W$.

Proof. i. First, note that since f is a linear map, we have $f(\mathbf{0}_\mathrm{V}) = \mathbf{0}_\mathrm{W}$.
$\boxed{\Longrightarrow}$ Let f be injective. Assume, by contradiction, that $\ker f$ is not trivial, i.e., there exists $\mathbf{v} \neq \mathbf{0}_\mathrm{V}$ such that $f(\mathbf{v}) = \mathbf{0}_\mathrm{W}$. Let $\mathbf{v}' \in V$. Then, using the linearity and also the injectivity of f, we find

$$\mathbf{0}_\mathrm{W} = f(\mathbf{v}) = f(\mathbf{v}-\mathbf{v}'+\mathbf{v}') = f(\mathbf{v}-\mathbf{v}')+f(\mathbf{v}') \Longleftrightarrow f(\mathbf{v}-\mathbf{v}') = -f(\mathbf{v}') = f(-\mathbf{v}') \Longrightarrow \mathbf{v}-\mathbf{v}' = -\mathbf{v}' \Longleftrightarrow \mathbf{v} = \mathbf{0}_\mathrm{V},$$

which is a contradiction. Hence, we must have $\ker f = \{\mathbf{0}_\mathrm{V}\}$.
$\boxed{\Longleftarrow}$ Let $\ker f = \{\mathbf{0}_\mathrm{V}\}$. Assume, by contradiction that f is not injective, i.e., there are two distinct vectors in V, such that $f(\mathbf{v}_1) = f(\mathbf{v}_2)$, with $\mathbf{v}_1 \neq \mathbf{v}_2$. Let $\mathbf{v} = \mathbf{v}_1 - \mathbf{v}_2 \neq \mathbf{0}_\mathrm{V}$. Then

$$f(\mathbf{v}) = f(\mathbf{v}_1 - \mathbf{v}_2) = f(\mathbf{v}_1) - f(\mathbf{v}_2) = \mathbf{0}_\mathrm{W} \Longrightarrow \mathbf{v} \in \ker f,$$

which is a contradiction. Hence, f is injective.

ii. First, note that by definition of image set, $\mathrm{im}\ f \subseteq W$.
$\boxed{\Longrightarrow}$ Let f be surjective. By definition, any $\mathbf{w} \in W$ satisfies $\mathbf{w} = f(\mathbf{v}) \in \mathrm{im}\ f$, so that $W \subseteq \mathrm{im}\ f$. Since $\mathrm{im}\ f \subseteq W$, there holds $W = \mathrm{im}\ f$.
$\boxed{\Longleftarrow}$ Let $\mathrm{im}\ f = W$. Let $\mathbf{w} \in W$. Then $\mathbf{w} \in \mathrm{im}\ f$, so that $\mathbf{w} = f(\mathbf{v})$ for some $\mathbf{v} \in V$. Since $\mathbf{w}$ is arbitrary, by definition, f is surjective. ■

Let us consider some standard examples of linear maps.

11.2 Examples

First, note two special linear maps that will arise in our later discussion.

Example 11.1 — Zero map. The **zero map** $o : V \to W$ is a linear map (check this) given by $o(\mathbf{v}) = \mathbf{0}_\mathrm{W}$ for all $\mathbf{v} \in V$. This means that $\ker(o) = V$ and $\mathrm{im}\ (o) = \{\mathbf{0}_\mathrm{W}\}$. The map is injective if and only if $V = \{\mathbf{0}_\mathrm{V}\}$ and surjective if and only $W = \{\mathbf{0}_\mathrm{W}\}$.

Example 11.2 — Identity map. The **identity map** $id : V \to V$ is a linear map (check this) given by $id(\mathbf{v}) = \mathbf{v}$, for all $\mathbf{v} \in V$. Note that the map is an endomorphism, as $W = V$. The kernel of id is trivial, while $\mathrm{im}\ id = V$. Hence, id is both injective and surjective.

Example 11.3 — Differentiation map. The **differentiation map** $D : \mathcal{P}_n(\mathbb{R}) \to \mathcal{P}_{n-1}(\mathbb{R})$ is a linear map given by $(Dp)(x) = p'(x)$. It is linear since, using the properties of derivatives,

$$(D(ap + bq))\,(x) = (ap(x) + bq(x))' = ap'(x) + bq'(x) = aDp(x) + bDq(x).$$

Its kernel is non-trivial, since the constant polynomial $p(x) = a$, for $a \in \mathbb{R}$, yields $Dp(x) = p'(x) = 0$. The codomain $\mathcal{P}_{n-1}(\mathbb{R})$ is chosen to ensure surjectivity.

Example 11.4 — Integration map. The **definite integration map** $I : C^0([-1, 1]) \to \mathbb{R}$ is a linear map given by $I(f) = \int_{-1}^{1} f(x)\mathrm{d}x$. It is linear since, using the properties of integrals,

$$I(af + bg) = \int_{-1}^{1} [af(x) + bg(x)]\,\mathrm{d}x = a\int_{-1}^{1} f(x)\mathrm{d}x + b\int_{-1}^{1} g(x)\mathrm{d}x = aI(f) + bI(g).$$

Its kernel is non-trivial, since $I(f) = 0$ for any even function f.

Example 11.5 — Matrix multiplication map. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be given by $f(\mathbf{x}) = A\mathbf{x}$ for $\mathbf{x} \in \mathbb{R}^n$. Then f is a linear map since, using the properties of matrix-vector multiplication,

$$f(a\mathbf{x} + b\mathbf{y}) = A(a\mathbf{x} + b\mathbf{y}) = aA\mathbf{x} + bA\mathbf{y} = af(\mathbf{x}) + bf(\mathbf{y}).$$

This transformation is fundamental and will be studied in a later lecture.

Example 11.6 — Coordinate map. Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ denote a basis for V. Let

$$\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \ldots + x_n\mathbf{v}_n.$$

Define the **coordinate map** $\varphi : V \mapsto \mathbb{F}^n$ via

$$\varphi(\mathbf{v}) = \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Then φ is a linear map. To see, this, let $\mathbf{v}, \mathbf{w} \in V$ have their representations in the basis B given below

$$\mathbf{v} = \sum_{j=1}^{n} a_j\mathbf{v}_j, \quad \mathbf{w} = \sum_{j=1}^{n} b_j\mathbf{v}_j.$$

Then

$$a\mathbf{v} + b\mathbf{w} = a\sum_{j=1}^{n} a_j\mathbf{v}_j + b\sum_{j=1}^{n} b_j\mathbf{v}_j = \sum_{j=1}^{n}(aa_j + bb_j)\mathbf{v}_j =: \sum_{j=1}^{n} c_j\mathbf{v}_j.$$

Hence,

$$\varphi(a\mathbf{v} + b\mathbf{w}) = \mathbf{c} = a\mathbf{x} + b\mathbf{y} = a\varphi(\mathbf{v}) + b\varphi(\mathbf{w}).$$

This map is bijective, so by Proposition 11.2, its kernel is trivial and im $\varphi = \mathbb{R}^n$.

LECTURE 12

Linear transformations (cont.)

12.1 Subspaces

It is natural to ask what happens when we try to map a subspace U of V: is $f(U)$ also a subspace of $f(V)$? Are there some obvious and/or some special cases? We consider these matters below. We first consider the image and the kernel of a linear map.

Proposition 12.1 Let $f : V \to W$ be a linear map. Then
i. $\operatorname{im} f$ is a subspace of W: $\operatorname{im} f \leq W$.
ii. $\ker f$ is a subspace of V: $\ker f \leq V$.

Proof. First, note that since f is a linear map, we have $f(\mathbf{0}_{\mathrm{V}}) = \mathbf{0}_{\mathrm{W}}$. Therefore, $\operatorname{im} f$ and $\ker f$ are non-empty subsets of W and V, respectively. This allows us to apply the subspace criterion 2 in both cases.
i. Let $\mathbf{w}_1, \mathbf{w}_2 \in \operatorname{im} f$; then there exist $\mathbf{v}_1, \mathbf{v}_2 \in V$ such that $\mathbf{w}_1 = f(\mathbf{v}_1), \mathbf{w}_2 = f(\mathbf{v}_2)$. Then, for any scalars $a_1, a_2 \in \mathbb{F}$, we have

$$a_1\mathbf{w}_1 + a_2\mathbf{w}_2 = a_1 f(\mathbf{v}_1) + a_2 f(\mathbf{v}_2) = f(a_1\mathbf{v}_1 + a_2\mathbf{v}_2) =: f(\mathbf{v}) \in \operatorname{im} f.$$

Hence, $\operatorname{im} f$ is a subspace of W: $\operatorname{im} f \leq W$.
ii. Let $\mathbf{v}_1, \mathbf{v}_2 \in \ker f$. Then, for any scalars $a_1, a_2 \in \mathbb{F}$, the vector $\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2$ is in $\ker f$ since

$$f(\mathbf{v}) = f(a_1\mathbf{v}_1 + a_2\mathbf{v}_2) = a_1 f(\mathbf{v}_1) + a_2 f(\mathbf{v}_2) = a_1\mathbf{0}_{\mathrm{W}} + a_2\mathbf{0}_{\mathrm{W}} = \mathbf{0}_{\mathrm{W}}.$$

Hence, $\ker f$ is a subspace of V: $\ker f \leq V$. ∎

Proposition 12.2 Let $U \leq V$. Then $f(U) \leq W$.

Proof. The proof is left as an exercise. ∎

12.2 Spans, bases, dimension

Given a spanning set S for a vector space V, by definition, im $f = f(V) = f(\text{span}S)$. The following result confirms that spanning sets are sufficient to describe the image of a map in the following sense.

Proposition 12.3 Let $f : V \to W$ be a linear map and let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ be a set of vectors in V. Then $f(\text{span}S) = \text{span}f(S)$.

Proof. Let $W \ni \mathbf{w}_i = f(\mathbf{v}_i), i = 1, 2, \dots, k$. With this notation, the result follows from the identity

$$f\left(\sum_{i=1}^{k} a_i \mathbf{v}_i)\right) = \sum_{i=1}^{k} a_i f(\mathbf{v}_i) = \sum_{i=1}^{k} a_i \mathbf{w}_i.$$

■

R This result holds, in particular, for the case where S is a basis for V.

We can establish similar or related results for linearly independent sets.

Proposition 12.4 Let $f : V \to W$ be a linear map with trivial kernel. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\} \in V$. Then S is a linearly independent set in V if and only if $f(S)$ is a linearly independent set in W.

Proof. We have

$$\mathbf{0}_\text{V} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k \Longleftrightarrow \mathbf{0}_\text{W} = a_1 f(\mathbf{v}_1) + a_2 f(\mathbf{v}_2) + \cdots + a_k f(\mathbf{v}_k).$$

Note that the reverse implication always holds as $f(\mathbf{0}_\text{V}) = \mathbf{0}_\text{W}$ by linearity of f, while the direct implication holds since the kernel of f is trivial, i.e., $\mathbf{0}_\text{W} = f(\mathbf{v})$ only if $\mathbf{v} = \mathbf{0}_\text{V}$. The result then follows from the above equivalence: a linear combination in V is trivial if and only if it is trivial in W. ■

Note that without the assumption on the kernel of f, we can only show the following.

Proposition 12.5 Let $f : V \to W$ be a linear map and let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$. If $f(S)$ is a linearly independent set in W, then S is a linearly independent set in V.

Corollary 12.6 Let $f : V \to W$ be a linear map with trivial kernel. Let S be a linearly independent set in V. Then dim $f(\text{span}S) = |S|$.

Proof. By Proposition 12.4, $f(S)$ is a linearly independent set in W and therefore a basis for the subspace U of W that it spans, namely $U = \text{span}f(S)$. By Proposition 12.3, $U = \text{span}f(S) = f(\text{span}S)$. Hence dim $f(\text{span}S) = |f(S)| = |S|$. ■

12.3 Rank and nullity

Definition 12.1 — Rank and nullity. Let $f : V \to W$ be a linear map.
The **rank** of f is the dimension of the image of f: rank f = dim im f.
The **nullity** of f is the dimension of the kernel of f: nullity f = dim ker$f$.

Example 12.1 If $f = o$ (the zero map), then im $o = \{\mathbf{0}_\text{W}\}$, so that rank $o = 0$. Since for all $\mathbf{v} \in V$, $o(\mathbf{v}) = \mathbf{0}_\text{W}$, nullity o = dim V. On the other hand, if $f = id$, then im $id = V$, so that rank id = dim W = dim V. Finally, since $id(\mathbf{0}_\text{V}) = \mathbf{0}_\text{V}$, the kernel of id is trivial and hence nullity $id = 0$.

The following result contains observations based on previous definitions.

Proposition 12.7 Let $f : V \to W$ be a linear map, where V is a finite dimensional vector space. Then

$$0 \le \text{rank } f \le \dim W, \qquad 0 \le \text{nullity } f \le \dim V.$$

We are now ready to prove the following fundamental result.

Theorem 12.8 — Rank-nullity formula. Let V be an $n-$dimensional vector space. Let $f : V \to W$ be a linear map. Then

$$\text{rank } f + \text{nullity } f = n.$$

Proof. Let B denote a basis set for V containing a basis B_1 for $\ker f$ (see Proposition 6.3). Denote by B_2 the complement of B_1 in B; then $B = \{B_1, B_2\}$, where, by construction,

- B_1 and B_2 are disjoint sets;
- B_2 is a linearly independent set.

Define $k := |B_1| = \dim \ker f = \text{nullity } f$ and $r := |B_2|$. With this notation,

$$n = \dim V = |B| = |B_1| + |B_2| = k + r = \text{nullity } f + r.$$

Claim: $r = \text{rank } f$. To see this, consider the linear map $\tilde{f} : \text{span} B_2 \to W$ defined via $\tilde{f}(\mathbf{v}) = f(\mathbf{v})$ for $\mathbf{v} \in \text{span} B_2$. Note that $f(B_2) = \tilde{f}(B_2)$. Since the kernel of $\tilde{f}$ is trivial due to the disjointness of B_1 and B_2, we can use Proposition 12.4 to deduce that $\tilde{f}(B_2)$ is a linearly independent set in W. Moreover, it is a spanning set for $\text{im } f$ since

$$\text{im } f = f(V) = f(\text{span} B) = \text{span} f(B) = \text{span}\{f(B_1), f(B_2)\} = \text{span}\{\mathbf{0}_{\mathrm{W}}, f(B_2)\} = \text{span} f(B_2) = \text{span} \tilde{f}(B_2),$$

where we used the result of Proposition 12.3. Hence, $\tilde{f}(B_2)$ is a basis for $\text{im } f$ and, by definition,

$$\text{rank } f = \dim \text{span} \tilde{f}(B_2) = \left| \tilde{f}(B_2) \right| = |B_2| = r$$

and the result follows. ■

We end this lecture with the following results on injectivity and surjectivity.

Proposition 12.9 Let V be an $n-$dimensional vector space. Let $f : V \to W$ be a linear map.
i. If $\dim V > \dim W$, then f is not injective.
ii. If $\dim V < \dim W$, then f is not surjective.

Proof. Both results are consequences of the rank-nullity formula.
i. Let $\dim V > \dim W$. Then f is not injective as the kernel of f is not trivial since

$$\dim \ker f = \text{nullity } f = n - \text{rank } f \ge n - \dim W > 0$$

ii. Let $\dim V < \dim W$. Then f is not surjective as $\text{im } f \ne W$ since

$$\dim \text{im } f = \text{rank } f = n - \text{nullity } f \le \dim V < \dim W.$$

■

LECTURE 13

Matrix representations

13.1 Matrix representations

By the rank-nullity formula, if the kernel of a map $f : V \to W$ is trivial, then the dimension of the image is the same as that of the domain. Since linearly independent sets in V are mapped to linearly independent sets in $f(V)$, the action of f on a basis of V will yield a basis for $f(V)$. Thus, the action of f on a vector $\mathbf{v} \in V$ can be reduced to the action of f on basis elements. Let us consider an example to see how this works.

Example 13.1 Let $f : \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_1(\mathbb{R})$ be given by

$$f(a_0 + a_1 x + a_2 x^2) = a_2 + (a_0 + a_1 + a_2)x.$$

Consider the following bases for $V = \mathcal{P}_2(\mathbb{R})$ and $W = \mathcal{P}_1(\mathbb{R})$, respectively,

$$B_V = \{p_1, p_2, p_3\} := \left\{1, 1+x, 1+x+x^2\right\}, \qquad B_W = \{q_1, q_2\} := \{1, x\}.$$

The action of f on each basis element of $\mathcal{P}_2$ is as follows:

$$\begin{cases} f(p_1) = f(1) & = x = 0 \cdot q_1 + 1 \cdot q_2, \\ f(p_2) = f(1+x) & = 2x = 0 \cdot q_1 + 2 \cdot q_2, \\ f(p_3) = f(1+x+x^2) & = 3x + 1 = 1 \cdot q_1 + 3 \cdot q_2. \end{cases}$$

Thus, we can establish the correspondence

$$p_1 \longleftrightarrow \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \xrightarrow{f} \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad p_2 \longleftrightarrow \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \xrightarrow{f} \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \quad p_3 \longleftrightarrow \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \xrightarrow{f} \begin{bmatrix} 1 \\ 3 \end{bmatrix},$$

where we used the convention that the coordinates in the bases B_V, B_W are displayed in column vectors.

We can therefore summarise the action of f on the basis of V as follows

$$[p_1, p_2, p_3] \longleftrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{f} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 2 & 3 \end{bmatrix}.$$

On the other hand, any $p \in \mathcal{P}_2(\mathbb{R})$ will be mapped by f as follows:

$$f(p) = f(c_1p_1 + c_2p_2 + c_3p_3) = c_1f(p_1) + c_2f(p_2) + c_3f(p_3)$$

$$= c_1x + c_2(2x) + c_3(3x + 1) = c_3q_1 + (c_1 + 2c_2 + 3c_3)q_2 = q.$$

Hence, the action of mapping p to q can be viewed as the action of mapping the coordinates of p to the coordinates of q via a rectangular matrix:

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \xrightarrow{\begin{bmatrix} 0 & 0 & 1 \\ 1 & 2 & 3 \end{bmatrix}} \begin{bmatrix} c_3 \\ c_1 + 2c_3 + 3c_3 \end{bmatrix},$$

or equivalently,

$$f(p) \longleftrightarrow \begin{bmatrix} 0 & 0 & 1 \\ 1 & 2 & 3 \end{bmatrix} * \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \longleftrightarrow \begin{bmatrix} c_3 \\ c_1 + 2c_3 + 3c_3 \end{bmatrix} \longleftrightarrow q.$$

The operation $*$ can be immediately seen as the matrix-vector product between the matrix associated with f and the coordinates of a vector in the domain of f. In fact, this example represents *the motivation for the definition of the matrix-product as we know it.* It is also the motivation for the definition of a matrix representation of a linear map – included below.

For simplicity of presentation, in the following we consider only vector spaces over $\mathbb{R}$.

Definition 13.1 Let V, W be finite-dimensional vector spaces over $\mathbb{R}$. Let $f : V \to W$ be a linear map and assume that

$$B_V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}, \qquad B_W = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$$

are basis sets for V and W, respectively. For $i = 1, 2, \ldots, n$, let

$$f(\mathbf{v}_i) = \sum_{j=1}^{m} a_{ji}\mathbf{w}_j.$$

The matrix $A := [a_{ji}] \in \mathbb{R}^{m\times n}$ is **the matrix representation of f relative to the bases B_V, B_W.**

R Note that the basis sets have to be considered as ordered sets in the above definition, as different orderings lead to permuted versions of the matrix A.

The above definition is validated by the following general result.

Proposition 13.1 Let $f \in \mathcal{L}(V, W)$ and let $A \in \mathbb{R}^{m\times n}$ be its matrix representation relative to bases

$$B_V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}, \qquad B_W = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}.$$

Let

$$\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n = \sum_{i=1}^{n} x_i\mathbf{v}_i \in V$$

and

$$f(\mathbf{v}) = y_1\mathbf{w}_1 + y_2\mathbf{w}_2 + \cdots + y_m\mathbf{w}_m = \sum_{j=1}^{m} y_j\mathbf{w}_j \in W.$$

Then $\mathbf{y} = A\mathbf{x}$, where $y_j = [\mathbf{y}]_j$ for $j = 1, 2, \ldots, m$ and $x_i = [\mathbf{x}]_i$ for $i = 1, 2, \ldots, n$.

Proof. We have

$$f(\mathbf{v}) = f(x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \ldots + x_n\mathbf{v}_n) = \sum_{j=1}^{n} x_j f(\mathbf{v}_j) = \sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij}\mathbf{w}_i = \sum_{i=1}^{m}\sum_{j=1}^{n} x_j a_{ij}\mathbf{w}_i = \sum_{i=1}^{m} y_i\mathbf{w}_i,$$

where, for $i = 1, 2, \ldots, m$,

$$y_i := \sum_{j=1}^{n} x_j a_{ij} = \sum_{j=1}^{n} a_{ij}x_j = [A\mathbf{x}]_i.$$

■

The above statement has the following counterpart.

Proposition 13.2 Let $A \in \mathbb{R}^{m\times n}$. Then A is the matrix representation of some linear map relative to any bases for any vector spaces of dimensions n and m, respectively.

Proof. Let V, W be arbitrary vector spaces of dimensions n and m, with arbitrary bases

$$B_V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}, \qquad B_W = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}.$$

Then any $\mathbf{v} \in V$ can be represented as

$$\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n = \sum_{j=1}^{n} x_j\mathbf{v}_j = \sum_{j=1}^{n} [\mathbf{x}]_j\mathbf{v}_j, \qquad [\mathbf{x}]_j := x_j \ (j = 1, \ldots, n).$$

Define $\mathbf{y} := A\mathbf{x}$, i.e.,

$$y_i := [\mathbf{y}]_i := [A\mathbf{x}]_i \ (i = 1, \ldots, m).$$

Since $\mathbf{v}$ is arbitrary in V, we can define a map $f : V \to W$ via

$$f(\mathbf{v}) := y_1\mathbf{w}_1 + y_2\mathbf{w}_2 + \cdots + y_m\mathbf{w}_m = \sum_{i=1}^{m} [\mathbf{y}]_i\mathbf{w}_i.$$

Claim: f is a linear map. To see this, let $\mathbf{v}$ be as above and choose another vector $\tilde{\mathbf{v}} \in V$ represented as

$$\tilde{\mathbf{v}} = \tilde{x}_1\mathbf{v}_1 + \tilde{x}_2\mathbf{v}_2 + \ldots + \tilde{x}_n\mathbf{v}_n = \sum_{j=1}^{n} \tilde{x}_j\mathbf{v}_j = \sum_{j=1}^{n} [\tilde{\mathbf{x}}]_j\mathbf{v}_j, \qquad [\tilde{\mathbf{x}}]_j := \tilde{x}_j \ (j = 1, \ldots, n).$$

Define $\tilde{\mathbf{y}} = A\tilde{\mathbf{x}}$, i.e.,

$$\tilde{y}_i := [\tilde{\mathbf{y}}]_i := [A\tilde{\mathbf{x}}]_i \ (i = 1, \ldots, m).$$

Then, by the above definition of f,

$$f(\tilde{\mathbf{v}}) = \tilde{y}_1\mathbf{w}_1 + \tilde{y}_2\mathbf{w}_2 + \cdots + \tilde{y}_m\mathbf{w}_m = \sum_{i=1}^{m} [\tilde{\mathbf{y}}]_i\mathbf{w}_i.$$

We have

$$af(\mathbf{v}) + bf(\tilde{\mathbf{v}}) = a\left(\sum_{i=1}^{m}[\mathbf{y}]_i\mathbf{w}_i\right) + b\left(\sum_{i=1}^{m}[\tilde{\mathbf{y}}]_i\mathbf{w}_i\right) = \sum_{i=1}^{m}[a\mathbf{y} + b\tilde{\mathbf{y}}]_i\mathbf{w}_i.$$

On the other hand,

$$a\mathbf{v} + b\tilde{\mathbf{v}} = a\left(\sum_{j=1}^{n}[\mathbf{x}]_j\mathbf{v}_j\right) + b\left(\sum_{j=1}^{n}[\tilde{\mathbf{x}}]_j\mathbf{v}_j\right) = \sum_{j=1}^{n}[a\mathbf{x} + b\tilde{\mathbf{x}}]_j\mathbf{v}_j.$$

Since $A(a\mathbf{x} + b\tilde{\mathbf{x}}) = a\mathbf{y} + b\tilde{\mathbf{y}}$, we find

$$f(a\mathbf{v} + b\tilde{\mathbf{v}}) = \sum_{i=1}^{m}[a\mathbf{y} + b\tilde{\mathbf{y}}]_i\mathbf{w}_i = af(\mathbf{v}) + bf(\tilde{\mathbf{v}}).$$

Therefore, f is a linear map. By definition, A is the matrix representation of f relative to B_V, B_W. ■

We end this section with the following uniqueness result.

Proposition 13.3 Let $f \in \mathcal{L}(V, W)$. Then its matrix representation relative to bases B_V, B_W is unique.

Proof. Assume that there are two distinct matrix representations A, A' relative to bases

$$B_V = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}, \qquad B_W = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m\}.$$

By definition, for all $j = 1, 2, \dots, n$,

$$f(\mathbf{v}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{w}_i = \sum_{i=1}^{m} a'_{ij}\mathbf{w}_i.$$

Taking the difference, we find

$$\mathbf{0}_W = \sum_{i=1}^{m} a_{ij}\mathbf{w}_i - \sum_{i=1}^{m} a'_{ij}\mathbf{w}_i = \sum_{i=1}^{m} \tilde{a}_{ij}\mathbf{w}_i,$$

where $\tilde{a}_{ij} = a_{ij} - a'_{ij}$. Therefore, for all $j = 1, 2, \dots, n$, the zero vector $\mathbf{0}_W$ is a linear combination of basis elements; by linear independence, we have $\tilde{a}_{ij} = 0$ for all i, or $a_{ij} = a'_{ij}$ for all i, j, i.e., $A = A'$, which is a contradiction. ■

Before we consider some examples, we include the following counterpart to Proposition 11.1 which stated that $\mathcal{L}(V, W)$ is a vector space.

Proposition 13.4 The set of matrices $\mathbb{R}^{m\times n}$ is a vector space over $\mathbb{R}$.

Proof. Exercise. ■

This result confirms that the set of matrix representations for linear maps inherit the properties and structures afforded by the set of linear maps.

13.2 Examples

Let us consider some examples of derivations of matrices associated with linear maps. In the following, we assume that the domain and codomain are equipped with basis sets

$$B_V = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}, \qquad B_W = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m\}.$$

Example 13.2 — Zero map. Let $o : V \to W$ be the zero map with $A \in \mathbb{R}^{m\times n}$ as its matrix representation. Then, for all $j = 1, \dots, n$,

$$o(\mathbf{v}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{w}_i = \mathbf{0}_W.$$

In other words, given any j, the zero vector $\mathbf{0}_W$ is a linear combination of basis elements; by linear independence, we have $a_{ij} = 0$ for all i. Hence A is the $m \times n$ zero matrix $A = O_{m,n}$.

Example 13.3 — Identity map. Let $id : V \to V$ be the identity map with $A \in \mathbb{R}^{m\times n}$ as its matrix representation relative to bases B_V and $B_W = B_V$. Then, for all $j = 1, \dots, n$,

$$id(\mathbf{v}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{v}_i = \mathbf{v}_j \iff \sum_{i=1}^{m} \tilde{a}_{ij}\mathbf{v}_i = \mathbf{0}_V, \quad \text{where } \tilde{a}_{ij} = \begin{cases} a_{ij} & i \neq j, \\ a_{ij} - 1 & i = j. \end{cases}$$

Hence, given any j, the zero vector $\mathbf{0}_V$ is a linear combination of basis elements; by linear independence, we have $\tilde{a}_{ij} = 0$ for all i, so that for all j

$$a_{ij} = \begin{cases} 0 & i \neq j, \\ 1 & i = j. \end{cases}$$

Therefore, $A = I_n \in \mathbb{R}^{n\times n}$ is the identity matrix.

The next example considers a restriction of the differentiation map to spaces of polynomials. This ensures that the domain and codomain are finite-dimensional vector spaces.

Example 13.4 — Differentiation map on $\mathcal{P}_n(\mathbb{R})$. Let $D : \mathcal{P}_n(\mathbb{R}) \to \mathcal{P}_{n-1}(\mathbb{R})$ be given by $(Dp)(x) = p'(x)$. Note that $V = P_n(\mathbb{R})$, $W = P_{n-1}(R)$, with $\dim V = n+1, \dim W = n$. Let

$$B_V = \left\{q_j = x^{j-1}, j = 1, 2, \dots, n+1\right\}, \qquad B_W = \left\{q_i = x^{i-1}, i = 1, 2, \dots, n\right\}.$$

Then the matrix representation $A \in \mathbb{R}^{n\times(n+1)}$ satisfies for all $j = 1, \dots, n+1$

$$D(q_j) = \sum_{i=1}^{n} a_{ij} q_i.$$

Let us consider the case $j = 1$ first:

$$j = 1 : 0 = D(q_1) = \sum_{i=1}^{n} a_{i1} q_i \Longrightarrow a_{i1} = 0, \quad i = 1, 2, \dots, n.$$

For any $j > 1$, matching the monomials on the left and right, we find

$$D(q^j) = (j-1)x^{j-2} = \sum_{i=1}^{n} a_{ij} x^{i-1} \iff a_{ij} = \begin{cases} 0 & j-2 \neq i-1 \\ j-1 & j-2 = i-1 \end{cases} \iff a_{ij} = \begin{cases} 0 & i \neq j-1, \\ j-1 & i = j-1. \end{cases}.$$

The resulting matrix is included below.

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & n \end{bmatrix} \in \mathbb{R}^{n\times(n+1)}.$$

Our next example considers the anti-derivative/integration map. As for the previous example, to ensure that the domain and codomain are finite-dimensional, we restrict them to polynomial spaces.

Example 13.5 — Integration map on $\mathcal{P}_n(\mathbb{R})$. Let $V = \mathcal{P}_n(\mathbb{R})$. Define $W = \mathcal{P}^0_{n+1}(\mathbb{R}) = \{p \in \mathcal{P}_{n+1} : p(0) = 0\}$ and note that $W \leq \mathcal{P}_{n+1}(\mathbb{R})$. Let

$$B_V = \left\{q_j = x^{j-1}, j = 1, 2, \ldots, n+1\right\}, \qquad B_W = \left\{q_i = x^i, i = 1, 2, \ldots, n+1\right\}$$

and note that $\dim V = \dim W = n+1$. Let $I\colon V \to W$ be given by

$$I(p)(x) = \int p(x)\mathrm{d}x.$$

The above choice of codomain ensures that the anti-derivative $I(p)$ will result in a polynomial with zero constant term. The matrix representation $A \in \mathbb{R}^{(n+1)\times(n+1)}$ satisfies for all $j = 1, \ldots, n+1$

$$I(q_j) = \sum_{i=1}^{n+1} a_{ij} q_i \iff \frac{1}{j}x^j = \sum_{i=1}^{n+1} a_{ij} x^i \iff a_{ij} = \begin{cases} 0 & i \neq j, \\ 1/j & i = j, \end{cases} \iff a_{ij} = \frac{1}{j}\delta_{ij}.$$

The resulting matrix is diagonal and is included below:

$$A = \begin{bmatrix} 1/1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 1/2 & 0 & \cdots & \cdots & \vdots \\ \vdots & 0 & 1/3 & 0 & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 1/n+1 \end{bmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}.$$

13.3 Commutative diagrams

Proposition 13.1 indicates that the action of f on a vector $\mathbf{v}$ is associated with a matrix-vector product $\mathbf{y} = A\mathbf{x}$ involving the coordinate vectors $\mathbf{x}, \mathbf{y}$ of $\mathbf{v}$ and $f(\mathbf{v})$, respectively. We can make this more precise by introducing the concept of **coordinate map** (see Lecture 6). We define the following linear bijections $\varphi_V : V \to \mathbb{R}^n, \varphi_W : W \to \mathbb{R}^m$, via

$$\varphi_V(\mathbf{v}) = \mathbf{x}, \qquad \varphi_W(\mathbf{w}) = \mathbf{y}.$$

We also define the map $\alpha : \mathbb{R}^n \to \mathbb{R}^n$ via $\alpha(\mathbf{x}) = A\mathbf{x}$. Then

$$\varphi_W(f(\mathbf{v})) = \mathbf{y} = A\mathbf{x} = \alpha(\varphi_V(\mathbf{v})) \iff \alpha \circ \varphi_V = \varphi_W \circ f.$$

This is represented in the diagram below, which is known as a **commutative diagram**:

$$\begin{array}{ccc} \mathbf{v} & \xrightarrow{\quad f \quad} & \mathbf{w} = f(\mathbf{v}) \\ \varphi_V \downarrow & & \downarrow \varphi_W \\ \mathbf{x} & \xrightarrow{\quad \alpha \quad} & \mathbf{y} = A\mathbf{x} \end{array}$$

The term commutative can be best understood when $W = V$: in this case applying the action of f in V followed by the coordinate map φ_V yields the same result as applying the coordinate map followed by the action of f (i.e., of A) in $\mathbb{R}^n$. More generally, following the two paths from $\mathbf{v}$ indicated by the arrows yields the same result: $\mathbf{y}$. Should the maps be invertible (see later), we would be able to draw the arrows also in the opposite directions, thus allowing us to establish one-to-one correspondences. For this reason, commutative diagrams are useful, as they enable to identify more complex (compositions) of actions of linear maps. We will be using this device later when we discuss change of basis.

LECTURE 14

Matrix representations (cont.)

Having a matrix representation for a linear map provides an excellent tool for both analyis and computation. We will see that the properties of linear maps are paralleled by those of matrices; in fact, we often investigate matrices in order to understand better the linear maps under consideration. In this lecture, we consider some of the concepts and properties associated with linear maps which are naturally inherited by their matrix representations.

14.1 Properties of matrix representations

We start by looking at the expression of injectivity and surjectivity of a linear map in the corresponding matrix representations.

Proposition 14.1 Let $f : V \to W$ be a linear map with matrix representation $A \in \mathbb{R}^{m\times n}$ relative to bases B_V, B_W. Let $\mathbf{v}$ have coordinates $x_i = [\mathbf{x}]_i, i = 1, 2, \dots, n$ with respect to B_V. Then

$$\mathbf{v} \in \ker f \iff \mathbf{x} \in \ker A.$$

Proof Consider the following generic bases for V and W:

$$B_V = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}, \qquad B_W = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m\}.$$

We have, using the definition of matrix representation,

$$\mathbf{v} \in \ker f \iff f(\mathbf{v}) = \mathbf{0}_W \iff f\left(\sum_{i=1}^{n} x_i \mathbf{v}_i\right) = \mathbf{0}_W \iff \sum_{i=1}^{n} x_i f(\mathbf{v}_i) = \mathbf{0}_W \iff \sum_{i=1}^{n} x_i \sum_{j=1}^{m} a_{ji}\mathbf{w}_j = \mathbf{0}_W$$

$$\iff \sum_{j=1}^{m}\left(\sum_{i=1}^{n} x_i a_{ji}\right)\mathbf{w}_j = \mathbf{0}_W \iff \sum_{i=1}^{n} x_i a_{ji} = 0 \iff [A\mathbf{x}]_j = 0 \iff A\mathbf{x} = \mathbf{0}_n \iff \mathbf{x} \in \ker A.$$

■

We immediately derive the following corollary.

Corollary 14.2 Let $f : V \to W$ be a linear map with matrix representation $A \in \mathbb{R}^{m\times n}$ relative to bases B_V, B_W. Let $\varphi_V : V \to \mathbb{R}^n$ denote the coordinate map $\varphi_V(\mathbf{v}) = \mathbf{x}$, where $x_i = [\mathbf{x}]_i$ are the coordinates of $\mathbf{v}$ in the basis B_V. Then $\varphi_V(\ker f) = \ker A$.

Consequently, we get the following result.

Proposition 14.3 Let $f\colon V \to W$ be a linear map with matrix representation $A \in \mathbb{R}^{m\times n}$ relative to bases B_V, B_W. Then

$$\dim \ker f = \dim \ker A.$$

Proof If $\ker f$ is trivial, by the above corollary, so is $\ker A$; therefore, their dimensions are equal (to zero). Assume therefore that $\ker f$ is non-trivial and let B_0 denote a basis set for it. Let $\varphi_V\colon V \to \mathbb{R}^n$ denote the coordinate map $\varphi_V(\mathbf{v}) = \mathbf{x}$, where $x_i = [\mathbf{x}]_i$ are the coordinates of $\mathbf{v}$ in the basis B_V. Since φ_V is linear, by Corollary 12.6 and also using the above corollary,

$$\dim \ker f = |B_0| = \dim \varphi_V(\operatorname{span} B_0) = \dim \varphi_V(\ker f) = \dim \ker A.$$ ■

R The above results allow us to establish when the kernel of a linear map f is trivial by computing the kernel of its matrix representation with respect to some bases. Moreover, if the latter is non-trivial, we can immediately use the coordinate map to find the non-trivial kernel of f. Note that the choice of bases is arbitrary: once B_V, B_W are chosen, the computations can proceed in the usual way.

Example 14.1 Recall the linear map from Lecture 12: $f\colon \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_1(\mathbb{R})$ given by

$$f(a_0 + a_1x + a_2x^2) = a_2 + (a_0 + a_1 + a_2)x.$$

Consider the following bases for $V = \mathcal{P}_2(\mathbb{R})$ and $W = \mathcal{P}_1(\mathbb{R})$, respectively,

$$B_V = \{p_1, p_2, p_3\} := \left\{1, 1 + x, 1 + x + x^2\right\}, \qquad B_W = \{q_1, q_2\} := \{1, x\}.$$

Recall that the matrix representation with respect to these bases was found to be

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 2 & 3 \end{bmatrix}.$$

To find $\ker A$ we solve $A\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \iff \begin{cases} x_3 & = 0 \\ x_1 + 2x_2 & = 0 \end{cases} \iff \ker A = \operatorname{span}\left\{ \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} \right\}.$$

Hence, $\dim \ker f = 1$, with $\ker f = \operatorname{span}\{\mathbf{v}_0\}$, where

$$\mathbf{v}_0 = 2 \cdot p_1 + (-1) \cdot p_2 + 0 \cdot p_3 = 1 - x.$$

Check:

$$f(1 - x) = 0 + (1 - 1 + 0)x = 0.$$

Note that we could have also computed the kernel of f directly:

$$f(p)=0 \iff a_2+(a_0+a_1+a_2)\mathbf{x}=0 \iff \begin{cases} a_2 & =0 \\ a_0+a_1 & =0 \end{cases} \iff \ker f=\operatorname{span}\{p_0:=1-x\}.$$

However, in practice, it is more straightforward to implement (on a computer) the computation of $\ker A$.

Proposition 14.4 Let $f\colon V\to W$ be a linear map with matrix representation $A\in\mathbb{R}^{m\times n}$ relative to bases B_V, B_W. Let $\mathbf{v}\in V$ and let $\mathbf{w}=f(\mathbf{v})$ have coordinates $y_i=[\mathbf{y}]_i$. Then

$$\mathbf{w}\in \text{im } f \iff \mathbf{y}\in \text{col } A.$$

Proof Consider the coordinate maps $\varphi_V\colon V\to\mathbb{R}^n$, $\varphi_W\colon W\to\mathbb{R}^m$ given by

$$\varphi_V(\mathbf{v})=\mathbf{x} \qquad \varphi_W(\mathbf{w})=\mathbf{y},$$

where x_i are the coordinates of $\mathbf{v}$ in the basis B_V and $y_i=[\mathbf{y}]_i$ are the coordinates of $\mathbf{w}$ in the basis B_W. Recall also the commutative diagram introduced in the previous lecture:

$$\begin{tikzcd} \mathbf{v} \arrow[r, "f"] \arrow[d, "\varphi_V"'] & \mathbf{w}=f(\mathbf{v}) \arrow[d, "\varphi_W"] \\ \mathbf{x} \arrow[r, "\alpha"] & \mathbf{y}=A\mathbf{x} \end{tikzcd}$$

where we defined the linear map $\alpha(\mathbf{x}):=A\mathbf{x}$. In particular, recall that we found $\alpha\circ\varphi_V=\varphi_W\circ f$, a relation that we will use below. We have

$$\mathbf{w}\in \operatorname{im} f \iff \mathbf{w}=f(\mathbf{v}) \iff \varphi_W(\mathbf{w})=\varphi_W(f(\mathbf{v})) \iff \mathbf{y}=(\varphi_W\circ f)(\mathbf{v})=(\alpha\circ\varphi_V)(\mathbf{v})=A\mathbf{x} \iff \mathbf{y}\in\ker A.$$

■

We immediately derive the following corollary.

Corollary 14.5 Let $f\colon V\to W$ be a linear map with matrix representation $A\in\mathbb{R}^{m\times n}$ relative to bases B_V, B_W. Let $\mathbf{v}\in V$ and let $\mathbf{w}=f(\mathbf{v})$ have coordinates $y_i=[\mathbf{y}]_i$. Then $\varphi_W(\text{im } f)=\text{col } A$.

Consequently, we get the following result.

Proposition 14.6 Let $f\colon V\to W$ be a linear map with matrix representation $A\in\mathbb{R}^{m\times n}$ relative to bases B_V, B_W. Then

$$\text{rank } f=\text{rank } A.$$

Proof Let B denote a basis set for im f. Since φ_W is linear, by Corollary 12.6 and also using the above corollary,

$$\text{rank } f=\dim \text{ im } f=|B|=\dim\, \varphi_W(\operatorname{span}B)=\dim\, \varphi_W(\text{im } f)=\dim \text{ col } A=\text{rank } A.$$

■

An immediate consequence of the above results and the rank-nullity formula is the following.

Corollary 14.7 Let $A \in \mathbb{R}^{m\times n}$. Then

$$\dim \ker A = n - \text{rank } A.$$

In particular, the kernel of A is trivial if and only if its columns are linearly independent.

Exercise 14.1 Let $A \in \mathbb{R}^{m\times n}$. Show that if $m < n$, then $\ker A$ is non-trivial.

We end with a discussion of matrix representations under composition.

14.2 Matrix representation of compositions of linear maps

Let us recall, as well as introduce, some definitions, results and notation for composition of functions.

Definition 14.1 — Composition. Let $f \in \mathcal{L}(V, W)$ and $g \in \mathcal{L}(U, V)$. The composition of f and g is a map $h : U \to W$ given by $h(\mathbf{v}) := (f \circ g)(\mathbf{v}) = f(g(\mathbf{v}))$.

Proposition 14.8 Let $f \in \mathcal{L}(V, W)$ and $g \in \mathcal{L}(U, V)$. Then $h = f \circ g$ satisfies the following properties:

i. h is injective/surjective/bijective if f, g are injective/surjective/bijective;
ii. h is a linear map: $h \in \mathcal{L}(U, W)$;
iii. $h \neq g \circ f$, in general.

Proof Exercise. ■

R The composition $\circ$ is a **bilinear map**[1] $\mathcal{B} \colon \mathcal{L}(V, W) \times \mathcal{L}(U, V) \to \mathcal{L}(U, W)$ given by $h := \mathcal{B}(f, g) = f \circ g$. Note that $\mathcal{B}(\cdot, \cdot)$ is not symmetric since $f \circ g \neq g \circ f$, in general.

A natural concept to consider in this lecture is the form of the matrix representation of h, given the matrix representations of f and g, all relative to the bases for their domain and codomain. The following result establishes this relationship.

Proposition 14.9 Let U, V, W be finite-dimensional vector spaces equipped with basis sets B_U, B_V, B_W, respectively. Let $f \colon U \to W$ and $g \colon V \to U$ be linear maps and consider $h \colon V \to W$ given by the composition of f and g: $h \colon = f \circ g$. Consider the following matrix representations:

- $A \in \mathbb{R}^{m\times \ell}$ for linear map f, relative to B_U, B_W;
- $B \in \mathbb{R}^{\ell\times n}$ for linear map g, relative to B_V, B_U;
- $C \in \mathbb{R}^{m\times n}$ for linear map h, relative to B_V, B_W.

Then $C = AB$.

[1] See Definition 7.1.

R The corresponding commutative diagram is included below.

$$\begin{tikzcd}
\mathbf{v} \arrow[r, "g"] \arrow[d, "\varphi_V"'] & \mathbf{u} = g(\mathbf{v}) \arrow[r, "f"] \arrow[d, "\varphi_U"] & \mathbf{w} = f(\mathbf{u}) = f(g(\mathbf{v})) \arrow[d, "\varphi_W"] \\
\mathbf{x} \arrow[r, "B"] & \mathbf{y} = B\mathbf{x} \arrow[r, "A"] & \mathbf{z} = A\mathbf{y} = AB\mathbf{x}
\end{tikzcd}$$

Proof Consider the following bases for V, U, W:

$$B_V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}, \qquad B_U = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_\ell\}, \qquad B_W = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}.$$

Let $\mathbf{v} \in V$. Then, using the definition of matrix representation and Proposition 13.1, the action of h on a vector $\mathbf{v} \in V$ is described by multiplication by C

$$\mathbf{v} = \sum_{i=1}^{n} x_i \mathbf{v}_i \longrightarrow h(\mathbf{v}) = \sum_{k=1}^{m} [C\mathbf{x}]_k \mathbf{w}_k =: \sum_{k=1}^{m} z_k \mathbf{w}_k.$$

On the other hand, the action of g on a vector $\mathbf{v} \in V$ is described by multiplication by B

$$\mathbf{v} = \sum_{i=1}^{n} x_i \mathbf{v}_i \longrightarrow \mathbf{u} := g(\mathbf{v}) = \sum_{j=1}^{\ell} [B\mathbf{x}]_j \mathbf{u}_j =: \sum_{j=1}^{\ell} y_j \mathbf{u}_j,$$

while the action of f on a vector $\mathbf{u} \in U$ is described by multiplication by A

$$\mathbf{u} = \sum_{j=1}^{\ell} y_j \mathbf{u}_j \longrightarrow f(\mathbf{u}) = \sum_{k=1}^{m} [A\mathbf{y}]_k \mathbf{w}_k = \sum_{k=1}^{m} [AB\mathbf{x}]_k \mathbf{w}_k.$$

But $f(\mathbf{u}) = f(g(\mathbf{v})) = (f \circ g)(\mathbf{v}) = h(\mathbf{v})$. Hence,

$$h(\mathbf{v}) = \sum_{k=1}^{m} [C\mathbf{x}]_k \mathbf{w}_k = \sum_{k=1}^{m} [AB\mathbf{x}]_k \mathbf{w}_k \Longrightarrow C = AB.$$

■

The proof assumes that we have defined the product of two matrices A and B in the usual way. However, the derivation of the above matrix representation is the *motivation for the definition of the product of two matrices as we know it.* To see this, let us re-trace the proof as follows.

$$\mathbf{v} = \sum_{i=1}^{n} x_i \mathbf{v}_i \longrightarrow h(\mathbf{v}) = \sum_{k=1}^{m} \sum_{i=1}^{n} c_{ki} x_i \mathbf{w}_k.$$

On the other hand,

$$\mathbf{v} = \sum_{i=1}^{n} x_i \mathbf{v}_i \longrightarrow \mathbf{u} := g(\mathbf{v}) = \sum_{j=1}^{\ell} \sum_{i=1}^{n} b_{ji} x_i \mathbf{u}_j =: \sum_{j=1}^{\ell} y_j \mathbf{u}_j$$

and

$$\mathbf{u} = \sum_{j=1}^{\ell} y_j \mathbf{u}_j \longrightarrow f(\mathbf{u}) = \sum_{k=1}^{m} \sum_{j=1}^{m} a_{kj} y_j \mathbf{w}_k = \sum_{k=1}^{m} \sum_{j=1}^{m} a_{kj} \sum_{i=1}^{n} b_{ji} x_i \mathbf{w}_k = \sum_{k=1}^{m} \sum_{i=1}^{n} \sum_{j=1}^{m} a_{kj} b_{ji} x_i \mathbf{w}_k.$$

Since the last expression is $h(\mathbf{v})$, we deduce, by comparing the two expressions obtained above, that

$$c_{ki} = \sum_{j=1}^{m} a_{kj} b_{ji},$$

which is the definition we have for the matrix-matrix product. In particular, note that implied in this expression is the usual requirement that the number of columns of A (the range for j) is equal to the number of rows of B. In the case of the composition $f \circ g$, this is satisfied since $A \in \mathbb{R}^{m\times\ell}$ and $B \in \mathbb{R}^{\ell\times n}$.

We end with the following counterpart for the above proposition.

Proposition 14.10 Let U, V, W be vector spaces with dimensions $\dim U = \ell, \dim V = n, \dim W = m$. Let matrices $A \in \mathbb{R}^{m\times\ell}, B \in \mathbb{R}^{\ell\times n}$ and $C \in \mathbb{R}^{m\times n}$ satisfy $C = AB$. Then there exist linear maps $f\colon U \to W,\ g\colon V \to U$ and $h\colon V \to W$ with respective matrix representations A, B, C relative to some bases of U, V, W such that $h := f \circ g$.

R These two results underline once more the correspondence between linear transformations and matrices, including some of the standard properties and operations that arise in the study of linear maps.

LECTURE 15

Change of basis

Different bases lead to different representations of vectors as well as matrices associated with a linear map. In this lecture, we examine each case in detail.

15.1 Change of coordinates

Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, $B' = \{\mathbf{v}'_1, \mathbf{v}'_2, \ldots, \mathbf{v}'_n\}$ denote bases for a vector space V. Then any vector $\mathbf{v} \in V$ can be represented in each of the two bases, with different corresponding coordinates:

$$\mathbf{v} = \sum_{j=1}^{n} x_j \mathbf{v}_j = \sum_{i=1}^{n} x'_i \mathbf{v}'_i.$$

It is useful to establish the connection between the two sets of coordinates; we can do this by first considering the connection between the elements of the two bases. To this end, we note that each basis element in B can be written in the basis B' as follows:

$$\mathbf{v}_j = \sum_{i=1}^{n} m_{ij} \mathbf{v}'_i.$$

Using the above representations of $\mathbf{v}$, we find

$$\mathbf{v} = \sum_{j=1}^{n} x_j \mathbf{v}_j = \sum_{j=1}^{n} x_j \sum_{i=1}^{n} m_{ij} \mathbf{v}'_i = \sum_{i=1}^{n} \sum_{j=1}^{n} x_j m_{ij} \mathbf{v}'_i = \sum_{i=1}^{n} x'_i \mathbf{v}'_i$$

so that

$$x'_i = \sum_{i=1}^{n} m_{ij} x_j \Longleftrightarrow \mathbf{x}' = M\mathbf{x}.$$

The matrix M is called the **transition matrix**, as it allows for the change of coordinates from basis B to basis B'. Note that this equivalence can be seen as a special case of the matrix representation of the identity map with respect to different bases, as indicated graphically below (see first remark in L.12):

$$\begin{array}{ccc} \mathbf{v} & \xrightarrow{id} & \mathbf{v} = id(\mathbf{v}) \\ \varphi \big\downarrow & & \big\downarrow \varphi' \\ \mathbf{x} & \xrightarrow{M} & \mathbf{x}' = M\mathbf{x} \end{array}$$

Note that $M \neq I_n$ unless $B' = B$, which corresponds to no change of basis or coordinates. Note also that in some situations we may have to be precise with regard to the notation for the transition matrix. Where it is not clear from the context, we will use notation such as $M_{BB'}$ or $\mathbf{x}_{B'} = M_{BB'}\mathbf{x}_B$.

Proposition 15.1 The transition matrix $M_{BB'}$ is invertible.

Proof. Consider the transition matrices from B to B' and from B' to B. We have

$$\mathbf{x}_{B'} = M_{BB'}\mathbf{x}_B, \qquad \mathbf{x}_B = M_{B'B}\mathbf{x}_{B'}.$$

Hence,

$$\mathbf{x}_B = M_{B'B}M_{BB'}\mathbf{x}_B, \qquad \mathbf{x}_{B'} = M_{BB'}M_{B'B}\mathbf{x}_{B'},$$

so that

$$M_{B'B}M_{BB'} = I = M_{BB'}M_{B'B}.$$

By definition, $M_{BB'}$ is invertible with inverse $M_{BB'}^{-1} = M_{B'B}$. ∎

Let us consider an example.

Example 15.1 Let

$$B = \left\{1, x, x^2, x^3\right\} =: \{p_i : i = 1, \ldots, 4\}, \qquad B' = \left\{1, x, x^2 - \frac{1}{3}, x^3 - \frac{3}{5}x\right\} =: \{q_i, i = 1, \ldots, 4\}.$$

The transition matrix $M = M_{BB'} \in \mathbb{R}^{4\times4}$ has entries m_{ij} given by

$$p_j = \sum_{i=1}^{4} m_{ij}q_i$$

We have

$$\begin{cases} q_1(x) = 1 = m_{11} + m_{21}x + m_{31}\left(x^2 - \frac{1}{3}\right) + m_{41}\left(x^3 - \frac{3}{5}x\right) \\ q_2(x) = x = m_{12} + m_{22}x + m_{32}\left(x^2 - \frac{1}{3}\right) + m_{42}\left(x^3 - \frac{3}{5}x\right) \\ q_3(x) = x^2 = m_{13} + m_{23}x + m_{33}\left(x^2 - \frac{1}{3}\right) + m_{43}\left(x^3 - \frac{3}{5}x\right) \\ q_4(x) = x^3 = m_{14} + m_{24}x + m_{34}\left(x^2 - \frac{1}{3}\right) + m_{44}\left(x^3 - \frac{3}{5}x\right) \end{cases} \implies \begin{cases} m_{11} = 1, m_{21} = m_{31} = m_{41} = 0 \\ m_{22} = 1, m_{12} = m_{32} = m_{42} = 0 \\ m_{33} = 1, m_{13} - \frac{1}{3}m_{33} = 0, m_{23} = m_{43} = 0 \\ m_{44} = 1, m_{24} - \frac{3}{5}m_{44} = 0, m_{14} = m_{34} = 0 \end{cases}$$

so that

$$M_{BB'} = \begin{bmatrix} 1 & 0 & 1/3 & 0 \\ 0 & 1 & 0 & 3/5 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note that an alternative is to compute the entries in the matrix $M_{B'B}$, which is the inverse of $M_{BB'}$:

$$\begin{cases} q_1(x) = 1 & = m_{11} + m_{21}x + m_{31}x^2 + m_{41}x^3 \\ q_2(x) = x & = m_{12} + m_{22}x + m_{32}x^2 + m_{42}x^3 \\ q_3(x) = x^2 - \frac{1}{3} & = m_{13} + m_{23}x + m_{33}x^2 + m_{43}x^3 \\ q_4(x) = x^3 - \frac{3}{5}x & = m_{14} + m_{24}x + m_{34}x^2 + m_{44}x^3 \end{cases} \Longrightarrow M_{B'B} = \begin{bmatrix} 1 & 0 & -1/3 & 0 \\ 0 & 1 & 0 & -3/5 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

so that

$$M_{BB'} = \begin{bmatrix} 1 & 0 & -1/3 & 0 \\ 0 & 1 & 0 & -3/5 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 1/3 & 0 \\ 0 & 1 & 0 & 3/5 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Motivation: the polynomials in the basis B' are Legendre polynomials, which are orthogonal in the inner product

$$\langle p, q \rangle = \int_{-1}^{1} p(x)q(x)\mathrm{d}x.$$

For this reason, they are preferred in many applications and therefore transition matrices need to be computed.

15.2 Change of matrix representation

Let us consider now the matrix representation of a linear map $f : V \mapsto W$ under changes of bases. We will investigate this using diagrams. We first recall the diagram for the matrix representation relative to bases B_V, B_W using the notation indicated above:

$$\begin{tikzcd} \mathbf{v} \arrow[r, "f"] \arrow[d, "\varphi_V"'] & \mathbf{w} = f(\mathbf{v}) \arrow[d, "\varphi_W"] \\ \mathbf{x}_V \arrow[r, "A_{VW}"] & \mathbf{y}_W = A_{VW}\mathbf{x}_V \end{tikzcd}$$

Assume now that we change the basis B_W to B'_W. The resulting diagram is a combination of the previous two diagrams:

$$\begin{tikzcd} \mathbf{v} \arrow[r, "f"] \arrow[d, "\varphi_V"'] & \mathbf{w} = f(\mathbf{v}) \arrow[r, "id_{\mathrm{W}}"] \arrow[d, "\varphi_W"] & \mathbf{w} = id_W(\mathbf{w}) \arrow[d, "\varphi'_W"] \\ \mathbf{x}_V \arrow[r, "A_{VW}"] & \mathbf{y}_W = A_{VW}\mathbf{x}_V \arrow[r, "M_{WW'}"] & \mathbf{y}_{W'} = M_{WW'}\mathbf{y}_W \end{tikzcd}$$

Hence

$$\mathbf{y}_{W'} = M_{WW'}A_{VW}\mathbf{x}_V,$$

so that, by definition, the matrix representation of f with respect to bases B_V, B'_W is

$$A_{VW'} = M_{WW'}A_{VW}.$$

We can derive a similar relation for the case where we change the basis B_V. First, consider the corresponding diagram:

$$\begin{tikzcd}
\mathbf{v} \arrow[r, "id_V"] \arrow[d, "\varphi_{V'}"'] & \mathbf{v} = id(\mathbf{v}) \arrow[r, "f"] \arrow[d, "\varphi_V"] & \mathbf{w} = f(\mathbf{v}) \arrow[d, "\varphi_W"] \\
\mathbf{x}_{V'} \arrow[r, "M_{V'V}"] & \mathbf{x}_V = M_{V'V}\mathbf{x}_{V'} \arrow[r, "A_{VW}"] & \mathbf{y}_W = A_{VW}\mathbf{x}_V
\end{tikzcd}$$

Thus, $\mathbf{y}_W = A_{VW}M_{V'V}\mathbf{x}_{V'}$ and since $\mathbf{w} = (id_V \circ f)(\mathbf{v}) = f(\mathbf{v})$, we conclude that the matrix representations satisfy $A_{V'W} = A_{VW}M_{V'V}$. The final diagram, corresponding to changing both bases, is included below.

$$\begin{tikzcd}
\mathbf{v} \arrow[r, "id_V"] \arrow[d, "\varphi_{V'}"'] & \mathbf{v} = id(\mathbf{v}) \arrow[r, "f"] \arrow[d, "\varphi_V"] & \mathbf{w} = f(\mathbf{v}) \arrow[r, "id_{\mathrm{W}}"] \arrow[d, "\varphi_W"] & \mathbf{w} = id_W(\mathbf{w}) \arrow[d, "\varphi'_W"] \\
\mathbf{x}_{V'} \arrow[r, "M_{V'V}"] & \mathbf{x}_V = M_{V'V}\mathbf{x}_{V'} \arrow[r, "A_{VW}"] & \mathbf{y}_W = A_{VW}\mathbf{x}_V \arrow[r, "M_{WW'}"] & \mathbf{y}_{W'} = M_{WW'}\mathbf{y}_W
\end{tikzcd}$$

We obtain $\mathbf{y}_{W'} = M_{WW'}A_{VW}M_{V'V}\mathbf{x}_{V'}$, so that

$$A_{V'W'} = M_{WW'}A_{VW}M_{V'V}.$$

Note that by swapping V with V' and W with W' we obtain

$$A_{VW} = M_{W'W}A_{V'W'}M_{VV'}.$$

This relation can also be obtained by using the invertibility of $M_{VV'}$ and $M_{WW'}$. This discussion is summarised in the following theorem.

Theorem 15.2 Let $f : V \mapsto W$ be a linear map between finite-dimensional vector spaces V and W. Let

- B_V, B'_V be bases for V, with corresponding transition matrix $M_{VV'}$;
- B_W, B'_W bases for W, with corresponding transition matrix $M_{WW'}$.

Then the matrix representations of f relative to all four basis combinations satisfy

$$A_{VW} = A_{V'W}M_{VV'} = M^{-1}_{WW'}A_{VW'} = M^{-1}_{WW'}A_{V'W'}M_{VV'}.$$

R If in the above theorem we let $f = id\colon V \to V$, then assuming $A_{VV} = I_n$, we find

$$I_n = M^{-1}_{VV'}A_{V'V'}M_{VV'} \implies A_{V'V'} = M_{VV'}I_nM^{-1}_{VV'} = I_n.$$

This confirms that $id : V \to V$ has I_n as matrix representation, irrespective of choice of basis.

Exercise 15.1 Let V, W be vector spaces with dimensions n, m, respectively. Show that the zero map $o : V \to W$ has matrix representation $O_{m,n} \in \mathbb{R}^{m\times n}$, irrespective of choice of basis.

The expression in Theorem 15.2 connecting A_{VW} and $A_{V'W'}$ suggests the following definition:

Definition 15.1 — Matrix equivalence. We say $A, B \in \mathbb{F}^{m\times n}$ are equivalent if $\exists M \in \mathbb{F}^{m\times m}, N \in \mathbb{F}^{n\times n}$, both invertible, such that

$$B = M^{-1}AN.$$

By Theorem 15.2, two matrices are equivalent if they are the matrix representations of the same linear map. [The converse also holds: two equivalent matrices are the representations of some linear map relative to some suitably chosen bases]. The consequence is that any properties inherited from the linear map will be shared by the matrix representations. The first one to note is the rank.

Proposition 15.3 Let $A, B \in \mathbb{F}^{m\times n}$ be equivalent. Then rank A = rank B.

By the rank-nullity formula, we also have that nullity A = nullity B. Another feature that both A and B share is discussed next.

15.3 Canonical forms

Recall that an **equivalence relation**, denoted by $\sim$ is a binary relation on a set S satisfying the following three properties:

i. symmetry: $\alpha \sim \alpha$ for all $\alpha \in S$;
ii reflexivity: if $\alpha \sim \beta$, then $\beta \sim \alpha$ for all $\alpha, \beta \in S$;
iii. transitivity: if $\alpha \sim \beta$ and $\beta \sim \gamma$, then $\alpha \sim \gamma$, for all $\alpha, \beta, \gamma \in S$.

Given an equivalence class, it is natural to ask if there is a representative element in this class. This could be identified through some special (simple) form, or properties. Ideally, the criteria for designating this special form should identify a unique element from the equivalence class. We will refer to this as a **canonical** or **normal** form. We consider this for the case of matrix equivalence.

Proposition 15.4 Matrix equivalence is an equivalence relation on $\mathbb{F}^{m\times n}$.

Proof. Exercise. ■

We conclude that the matrix representations of a linear map belong to the same equivalence class. A candidate for a normal form is a diagonal matrix, due to its simplicity.

Proposition 15.5 Let V, W be vector spaces with dimensions n and m, respectively. Let $f \in \mathcal{L}(V, W)$ have rank r. Then there are bases B_V and B_W relative to which f has matrix representation

$$A_{VW} = \begin{bmatrix} I_r & O_{r,n-r} \\ O_{m-r,r} & O_{m-r,n-r} \end{bmatrix}.$$

Proof. The proof uses the results and notation included in the proof of Theorem 12.8 (the rank-nullity formula). Let $\{B_1, B_2\}$ be a basis for V, where

- $B_1 = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{n-r}\}$ is a basis for $\ker f$;
- $B_2 = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ is linearly independent.

By the proof of Theorem 12.8, the set $\tilde{B}_1 = \{\mathbf{w}_1 := f(\mathbf{v}_1), \mathbf{w}_2 := f(\mathbf{v}_2), \cdots, \mathbf{w}_r := f(\mathbf{v}_r)\}$ is a basis for $\operatorname{im} f$. Moreover, by Proposition 6.3, $\tilde{B}_1$ is contained in some basis B_W for W, say, $B_W = \left\{\tilde{B}_1, \tilde{B}_2\right\}$ for some set $\tilde{B}_2$. Define $\mathbf{v}_{i+r} := \mathbf{z}_i$ for $i = 1, 2, \dots, n-r$. Let

$$B_V := \{B_2, B_1\} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}.$$

With this choice of bases for V and W, we find

$$f(\mathbf{v}_i) = \begin{cases} \mathbf{w}_i & 1 \le i \le r, \\ \mathbf{0}_W & r+1 \le i \le n, \end{cases}$$

and the result follows. ■

We can explicitly see that the matrix A_{VW} has rank r, since it only have r linearly independent columns. Thus, the above result indicates that any matrix representation of f, has rank r, where $r = \operatorname{rank} f$.

Definition 15.2 The **canonical form** of the matrix representation of a linear map with rank r is the matrix

$$A^{\text{can}} = \begin{bmatrix} I_r & \\ & \end{bmatrix}.$$

LECTURE 16

Isomorphisms

The linear transformations studied so far are generally referred to as homomorphisms. They satisfy a linearity condition that ensures that their domain and range have a similar vector space structure: their elements behave similarly under vector addition and scalar-vector multiplication. However, they need not be in a one-to-one correspondence[1]. When this is the case, the domain can be 'identified' with the image. This type of mapping is called an isomorphism. We have already come across one such map: the coordinate mapping. Recall that given any $\mathbf{v} \in V$ expanded in a basis B_V as $\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$, we can define the mapping

$$\mathbf{v} \mapsto \mathbf{x} \in \mathbb{F}^n.$$

This linear map establishes a one-to-one correspondence between elements in a generic vector space V and $\mathbb{F}^n$. In this lecture, we outline some of the key results and properties associated with such maps. We start with invertibility.

16.1 Invertible maps

Let $f\colon V \to W$ be a linear map and let $\mathbf{v} \in V$. Recall that the **image** of $\mathbf{v}$ under f is a vector $\mathbf{w} \in W$ given by $\mathbf{w} = f(\mathbf{v})$. Indeed, we have defined the set of such images as the image set $f(V)$ (or im f):

$$f(V) := \{\mathbf{w} \in W : \mathbf{w} = f(\mathbf{v}), \mathbf{v} \in V\} \subseteq W.$$

Assume now that f is one-to-one (injective). Then **only one** vector in the domain of f is mapped to one vector in the image of f. Given this uniqueness, we can choose to define it as follows.

Definition 16.1 — Pre-image. Let $f \in \mathcal{L}(V, W)$ be injective and let $\mathbf{v} \in V$ and $\mathbf{w} = f(\mathbf{v}) \in \operatorname{im} f$. The vector $\mathbf{v}$ is called the **pre-image** or **inverse image** of $\mathbf{w}$ under f and is denoted by $\mathbf{v} = f^{-1}(\mathbf{w})$.

Not all $\mathbf{w} \in W$ are guaranteed to have an inverse image under f, unless f is surjective. On the other hand, by definition of image set, all $\mathbf{w} \in \operatorname{im} f$ have a pre-image when f is injective. Hence, we can extend/apply the notation f^{-1} to all elements of im f. Thus, we can view f^{-1} as a map with domain im f and codomain

[1] One-to-one correspondence means bijection; one-to-one mapping means injection.

V: f^{-1}: im $f \to V$. Moreover, note that any pre-image $\mathbf{v}$ of $\mathbf{w} \in \text{im } f = f(V)$, satisfies

$$\mathbf{v} = f^{-1}(\mathbf{w}) = f^{-1}(f(\mathbf{v})),$$

while any $\mathbf{w} \in \text{im } f$ satisfies

$$\mathbf{w} = f(\mathbf{v}) = f(f^{-1}(\mathbf{w})).$$

In other words, $f \circ f^{-1} = id_W$ and $f^{-1} \circ f = id_V$. This discussion justifies the following definition.

Definition 16.2 — Invertible map. Let $f\colon V \to W$ be a linear map. We say f is **invertible** if there exists a linear map $g\colon W \to V$ such that

$$g \circ f = id_V \quad \text{and} \quad f \circ g = id_W.$$

The map g is denoted by $g = f^{-1}$ and is called the **inverse** (map) of f.

Proposition 16.1 The inverse of a linear map $f\colon V \to W$ is unique, if it exists.

Proof. Let $g, h\colon W \to V$ be inverses of f. Then $h = h \circ id_W = h \circ (f \circ g) = (h \circ f) \circ g = id_V \circ g = g$. ■

Proposition 16.2 Let $f\colon V \to W$ be a linear map. Then f is invertible if and only if it is bijective.

Proof. Let f be a linear map.
$\boxed{\Longrightarrow}$ Assume f is invertible. Let $f(\mathbf{v}_1) = f(\mathbf{v}_2)$. Then $\mathbf{v}_1 = f^{-1}(f(\mathbf{v}_1) = f^{-1}(f(\mathbf{v}_2)) = \mathbf{v}_2$, so that f is injective. Let now $\mathbf{w} \in W$. Then there exists $\mathbf{v} = f^{-1}(\mathbf{w})$ such that $f(\mathbf{v}) = f(f^{-1}(\mathbf{w})) = \mathbf{w}$, so that f is surjective. Therefore f is bijective.
$\boxed{\Longleftarrow}$ Assume f is bijective. We show that (i) there exists a map $g\colon W \to V$ such that $g \circ f = id_V$ and $f \circ g = id_W$ and (ii) that g is linear. This is essentially the argument that motivated the definition of invertibility. Since f is surjective, every $\mathbf{w} \in W$ will have a corresponding pre-image $\mathbf{v} \in V$, which by injectivity is unique. Therefore, the map $g\colon W \to V$ given by $g(\mathbf{w}) = \mathbf{v}$ is well-defined. As a pre-image, $\mathbf{v}$ also satisfies $f(\mathbf{v}) = \mathbf{w}$. Hence,

$$\mathbf{w} = f(\mathbf{v}) = f(g(\mathbf{w})), \qquad \mathbf{v} = g(\mathbf{w}) = g(f(\mathbf{v}))$$

so that

$$f \circ g = id_W, \qquad g \circ f = id_V.$$

Finally, let $\mathbf{v}_1 = g(\mathbf{w}_1), \mathbf{v}_2 = g(\mathbf{w}_2)$. Then g is a linear map since

$$f(a\mathbf{v}_1 + b\mathbf{v}_2) = f(ag(\mathbf{w}_1) + bg(\mathbf{w}_2)) = af(g(\mathbf{w}_1)) + bf(g(\mathbf{w}_2)) = a\mathbf{w}_1 + b\mathbf{w}_2,$$

by linearity of f, so that

$$g(a\mathbf{w}_1 + b\mathbf{w}_2) = g(f(a\mathbf{v}_1 + b\mathbf{v}_2)) = a\mathbf{v}_1 + b\mathbf{v}_2 = ag(\mathbf{w}_1) + ag(\mathbf{w}_2).$$

■

Proposition 16.3 Let $f\colon V \to W$ be a linear map. Then f is invertible if and only if its matrix representation is invertible.

Proof. Let $f\colon V \to W$ be a linear map and let $n = \dim V$.
$\boxed{\Longrightarrow}$ Assume f is invertible. Then it is bijective and therefore $\ker f$ is trivial and $\operatorname{im} f = W$. By the rank-nullity formula, $\dim W = n$. Moreover, by definition of invertible maps, there exists a linear map such that

$$g \circ f = id_V \quad \text{and} \quad f \circ g = id_W.$$

By Proposition 14.9, the $n \times n$ matrix representations of f and g satisfy

$$A_g A_f = I_n \quad \text{and} \quad A_f A_g = I_n.$$

Hence, by the definition of matrix inverse, A_f has an inverse and is therefore invertible.
$\boxed{\Longleftarrow}$ Assume the matrix representation of f is a square matrix A which is invertible. Then there exists a matrix B such that

$$AB = I_n \quad \text{and} \quad BA = I_n.$$

By Proposition 14.10, the above relations correspond to compositions of linear maps f and g with matrix representations A and B, respectively in some bases for V and W. More precisely,

$$g \circ f = id_V \quad \text{and} \quad f \circ g = id_W.$$

By definition, f is an invertible map. ■

The corresponding commutative diagram is included below.

$$\begin{tikzcd}
\mathbf{v} \arrow[r, shift left, "f"] \arrow[d, shift right, "\varphi_V"'] & \mathbf{w} = f(\mathbf{v}) \arrow[l, shift left, "f^{-1}"] \arrow[d, shift left, "\varphi_W"] \\
\mathbf{x} \arrow[r, shift left, "A"] \arrow[u, shift right, "\varphi_V^{-1}"'] & \mathbf{y} = A\mathbf{x} \arrow[l, shift left, "A^{-1}"] \arrow[u, shift left, "\varphi_W^{-1}"]
\end{tikzcd}$$

16.2 Isomorphic spaces

Definition 16.3 A linear bijection $f\colon V \to W$ is called an **isomorphism**. In this case, the spaces V and W are said to be **isomorphic**; we write $V \cong W$.

Proposition 16.4 Let $f\colon V \to W$ be an isomorphism. Then $f^{-1}\colon W \to V$ is an isomorphism.

Proof. Since f is bijective, it is invertible, so the inverse map f^{-1} is linear. By definition, f^{-1} is invertible, therefore bijective, so an isomorphism. ■

Proposition 16.5 Let $f\colon V \to W$ be a linear map and let B_V be a basis for V. Then f is an isomorphism if and only if $f(B_V)$ is a basis for W.

Proof. Let $f\colon V \to W$ be a linear map and let $B_V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for V.
$\boxed{\Longrightarrow}$ Assume f is an isomorphism. Since f is injective, its kernel is trivial. Since f is surjective, $\operatorname{im} f = W$. By Proposition 12.4, since B_V is a linearly independent set, $f(B_V)$ is also linearly independent. Moreover, since f is linear, by Proposition 12.3, $\operatorname{span} f(B_V) = f(\operatorname{span} B_V) = f(V) = \operatorname{im} f = W$. Hence, $f(B_V)$ is a linearly independent spanning set for W, i.e., a basis for W.

$\boxed{\Longleftarrow}$ Assume that $f(B_V)$ is a basis for W. By Proposition 12.3, since f is linear, $\text{span} f(B_V) = f(\text{span} B_V)$, i.e., $W = \text{im } f$, so that f is surjective. To show f is injective, we show that $\ker f$ is trivial. Let $\mathbf{w}_i = f(\mathbf{v}_i)$ for $i = 1, 2, \ldots, n$ and note that $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ is a basis for W. Let $\mathbf{v} \in \ker f$. Then

$$\mathbf{0}_W = f(\mathbf{v}) = f\left(\sum_{i=1}^{n} x_i \mathbf{v}_i\right) = \sum_{i=1}^{n} x_i f(\mathbf{v}_i) = \sum_{i=1}^{n} x_i \mathbf{w}_i.$$

But this implies that $x_i = 0$ for all i, since $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ is a basis for W. Hence $\mathbf{v} = \mathbf{0}_V$ and the kernel is trivial. Thus, f is a bijection, which is also linear, and therefore an isomorphism. ■

Why are we interested in isomorphisms? It is because they allow for results or propertise derived for one space to also hold for the other space, as the following results shows.

Proposition 16.6 Let $f\colon V \to W$ be an isomorphism. Let $S \subseteq V$. Then

i. S is a spanning set for V if and only if $f(S)$ is a spanning set for W.
ii. S is linearly independent if and only if $f(S)$ is.
iii. S is a basis for V if and only if $f(S)$ is a basis for W.
iv. U is a subspace of V if and only if $f(U)$ is a subspace of W with same dimension.

Proof. Exercise. ■

The last item in the above proposition is related in some sense to the following key result.

Theorem 16.7 Let V, W be finite-dimensional vector spaces. Then

$$V \cong W \quad \text{if and only if} \quad \dim V = \dim W.$$

Proof. Let $f\colon V \to W$ be a linear map.
$\boxed{\Longrightarrow}$ Let $V \cong W$, with f a bijection. Since f is injective, $\ker f = \{\mathbf{0}_V\}$. Since f is surjective, $\text{im } f = W$. By the rank-nullity formula,

$$\dim V = \text{nullity } f + \text{rank } f = 0 + \dim W.$$

$\boxed{\Longleftarrow}$ Let $\dim V = \dim W = n$. Let $B_V = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}, B_W = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be bases for V and W, respectively. Define the map $f\colon V \to W$ via

$$f(\mathbf{v}) = f(x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n) = x_1\mathbf{w}_1 + x_2\mathbf{w}_2 + \cdots + x_n\mathbf{w}_n =: \mathbf{w}.$$

Then

- f is surjective, as for any $\mathbf{w} \in W$ with coordinates x_i in the basis B_W, there exists $\mathbf{v}$ having the same coordinates in the basis B_V.
- f is injective since

$$f(\mathbf{v}) = f(\mathbf{u}) \Longleftrightarrow \mathbf{w} = \mathbf{z} \Longleftrightarrow \sum_{i=1}^{n} x_i\mathbf{w}_i = \sum_{i=1}^{n} z_i\mathbf{w}_i \Longleftrightarrow \sum_{i=1}^{n}(x_i - z_i)\mathbf{w}_i = \mathbf{0} \Longleftrightarrow x_i = z_i \Longleftrightarrow \mathbf{v} = \mathbf{u}.$$

  Note that we used the linear independence of B_W to deduce that $x_i = z_i$ for all i.
- f is linear (show this!).

Hence f is a linear bijection and therefore $V \cong W$. ■

R The isomorphism f described in the above theorem has matrix representation $A = I_n$:

$$\begin{tikzcd} \mathbf{v} \arrow[r, "f"] \arrow[d, "\varphi_V"'] & \mathbf{w} = f(\mathbf{v}) \arrow[d, "\varphi_W"] \\ \mathbf{x} \arrow[r, "A = I_n"] & \mathbf{x} \end{tikzcd}$$

Theorem 16.7 allows us to establish quickly which pairs of spaces are not isomorphic, simply by examining the dimensions of the spaces.

An important consequence of Theorem 16.7 is that we can view all n-dimensional real vector spaces as being essentially the same. In particular, they are all isomorphic to $\mathbb{R}^n$.

Corollary 16.8 Let $V(\mathbb{R})$ be an n-dimensional vector space. Then $V \cong \mathbb{R}^n$.

We cannot stress enough the importance of this result: since every finite-dimensional space is isomorphic to $\mathbb{R}^n$, we can restrict our study to this specific vector space, as properties and results obtained for $\mathbb{R}^n$ will be paralleled in any other vector space of dimension n.

R The more general result is $V(\mathbb{F}) \cong \mathbb{F}^n$, for any n-dimensional V.

The above descriptions allow us to establish a criterion for identifying isomorphisms.

Proposition 16.9 — Isomorphism criterion. Let $f: V \to W$ be a linear map on a finite-dimensional vector space V. Then f is an isomorphism if any only $\ker f = \{\mathbf{0}_V\}$ and $\dim V = \dim W$.

Example 16.1 — Coordinate map. Let V be an n-dimensional vector space over $\mathbb{R}$ with basis set $B_V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, v_n\}$ and consider the coordinate map $\varphi_V : V \to \mathbb{R}^n$ given by

$$\varphi_V(\mathbf{v}) = \mathbf{x},$$

where $\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$. We can readily check the isomorphism criterion:

- $\ker\varphi_V = \{\mathbf{0}_V\}$, since any $\mathbf{v} \in \ker\varphi_V$ satisfies

  $$\mathbf{0}_n = \varphi_V(\mathbf{v}) = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n \Longrightarrow x_1 = x_2 = \cdots = x_n = 0,$$

  since the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is linearly independent. Hence, any $\mathbf{v} \in \ker\varphi_V$ satisfies $\mathbf{v} = \mathbf{0}_V$.
- $\dim \mathbb{R}^n = \dim V$.

Exercise 16.1 Let $V(\mathbb{F}), W(\mathbb{F})$ have bases B_V, B_W and dimensions n, m, respectively. Define the map $m: \mathcal{L}(V, W) \to \mathbb{F}^{m\times n}$ that associates with a linear map f its matrix representation relative to B_V, B_W:

$$m(f) = A_{VW}.$$

Use the isomorphism criterion to check that m is an isomorphism and therefore that $\mathcal{L}(V, W) \cong \mathbb{F}^{m\times n}$.

Example 16.2 — Counter-example. Let $V = \mathbb{C}, W = \mathbb{R}^2$. For any $z \in \mathbb{C}$ we can define a bijection

$$f(z) = f(a + ib) = \begin{bmatrix} a \\ b \end{bmatrix}.$$

> However, V and W are not isomorphic since f does not satisfy the definition of linear map: recall that by definition V is a vector space $V(\mathbb{F})$ and W is a vector space $W(\mathbb{F})$. However, our vector spaces do not share the same field: $\mathbb{C}$ means $\mathbb{C}(\mathbb{C})$, while $\mathbb{R}^2$ means $\mathbb{R}^2(\mathbb{R})$. Hence, the linearity property is undefined in this case.

16.3 Matrix representations

Given the above isomorphism criterion (and also the discussion prior to it), it is evident that the matrix representation of an isomorphism f is square and invertible. This excludes the zero matrix, which means that while the set of isomorphisms is a subset of $\mathcal{L}(V, W)$, it cannot be a subspace. The set of square and invertible matrices forms a (non-Abelian) group under matrix multiplication; however, this property does not extend to the set of isomorphisms equipped with the operation of composition, since if fails to satisfy the closure property. Our next topic, endomorphisms, i.e., maps in $\mathcal{L}(V, V)$, will allow us to establish a group structure to the subset of bijections in $\mathcal{L}(V, V)$.

LECTURE 17

Endomorphisms

In this lecture we consider linear maps from a vector space to itself: $f\colon V \to V$, i.e., $f \in \mathcal{L}(V, V)$. This is a rich topic, which includes the concept of eigenvalues which we will study in some depth. As before, we will assume that the vector space $V = V(\mathbb{F})$ is finite-dimensional, with $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$.

17.1 Endomorphisms

Definition 17.1 A linear map $f\colon V \to V$ from a vector space V to itself is called an **endomorphism**.

When f is an endomorphism, we use the simplified notation $f \in \mathcal{L}(V)$, although occasionally we may also use the standard notation $f \in \mathcal{L}(V, V)$. Another notation in use is $\mathrm{End}(V)$.

What is special about the domain and codomain being equal? Here are some observations.

Proposition 17.1 $\mathcal{L}(V)$ is a subspace of $\mathcal{L}(V, W)$: $\mathcal{L}(V) \leq \mathcal{L}(V, W)$.

Proof. Exercise. ■

Proposition 17.2 Let $V(\mathbb{F})$ be a non-trivial n-dimensional vector space. Then $\mathcal{L}(V) \cong \mathbb{F}^{n\times n}$.

Proof. Use Exercise 16.1. ■

This result confirms that we can associate any endomorphism f with some square matrix. We will be able to say more in the case when f is invertible (see later).

Another property that is specific to endomorphisms is that $\mathcal{L}(V)$ affords **composition as a set operation**, which was not the case for $\mathcal{L}(V, W)$. In particular, we note that any map $f \in \mathcal{L}(V)$ can be composed with itself. Moreover, $\mathcal{L}(V)$ satisfies the following properties when equipped with $\circ$:

- closure: $f \circ g \in \mathcal{L}(V)$;
- associativity: $f \circ (g \circ h) = (f \circ g) \circ h$;
- existence of identity: $f \circ id_V = id_V \circ f = f$;
- distributivity of composition over addition: $f \circ (g + h) = f \circ g + f \circ h$;

- distributivity of addition over composition: $(f+g)\circ h = f\circ h + g\circ h$.

However, composition is not commutative; therefore, $\mathcal{L}(V)$ is a **non-commutative ring**, when equipped with the additional operation of composition. Note also that not every $f \in \mathcal{L}(V)$ has an inverse with respect to $\circ$ (a property which would have ensured that it is a group when equipped with $\circ$).

Another special property of endomorphisms is associated with a change of bases. Let $f\colon V \to V$ and let B, B' denote bases for V. Then the matrix representations of f in the two bases satisfy

$$A_{V'V'} = M_{VV'} A_{VV} M_{VV'}^{-1}.$$

This is an important relation between matrix representations: we record this in a new definition.

Definition 17.2 — Similar matrices. Matrices $A, B \in \mathbb{F}^{n\times n}$ are said to be similar if there exists an invertible matrix $M \in \mathbb{F}^{n\times n}$ such that

$$B = M^{-1}AM.$$

Similar matrices share certain properties: they have the same rank, determinant, eigenvalues, to name but a few. Invariance under similarity has important practical and theoretical implications. This will be discussed later. At this stage, we note the following fact.

Proposition 17.3 Matrix similarity is an equivalence relation on $\mathbb{F}^{n\times n}$.

Proof. Exercise. ∎

Let us consider some examples; we include in the discussion the corresponding matrix representations.

Example 17.1 Let us revisit a previous example: the differentiation map. However, this time we define it as an endomorphism: $D\colon V \to V$, with $V = \mathcal{P}_n(\mathbb{R})$ and $(Dp)(x) = p'(x)$. Let $B_V = \left\{q_j = x^{j-1}, j = 1,2,\ldots,n+1\right\}$. Then the matrix representation $A \in \mathbb{R}^{(n+1)\times(n+1)}$ satisfies for all $j = 1,\ldots,n+1$

$$D(q_j) = \sum_{i=1}^{n+1} a_{ij} q_i.$$

Let us consider the case $j = 1$ first:

$$j = 1 : 0 = D(q_1) = \sum_{i=1}^{n+1} a_{i1} q_i \Longrightarrow a_{i1} = 0, \quad i = 1,2,\ldots,n+1.$$

For any $j > 1$, matching the monomials on the left and right, we find

$$D(q^j) = (j-1)x^{j-2} = \sum_{i=1}^{n+1} a_{ij} x^{i-1} \Longleftrightarrow a_{ij} = \begin{cases} 0 & j-2 \neq i-1 \\ j-1 & j-2 = i-1 \end{cases} \Longleftrightarrow a_{ij} = \begin{cases} 0 & i \neq j-1, \\ j-1 & i = j-1. \end{cases}.$$

The resulting matrix is included below.

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & n \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}.$$

Note that A is singular. Moreover, note that $A^{n+1} = O_{n,n}$; for example, when $n = 3$ we have

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad A^2 = \begin{bmatrix} 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad A^3 = \begin{bmatrix} 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad A^4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

This is an example of a **nilpotent matrix**. Note that differentiating 4 times a degree 3 polynomial yields the zero polynomial, so the repeated action of the map D is paralleled by that of A.

An example that involves the composition of maps is included below.

Example 17.2 — Projection map. The projection map is a map $\pi\colon V \to V$ defined via $\pi^2 = \pi$. Note that V is a generic vector space, without an inner product defined on it (i.e., there is no concept of orthogonality implied in the definition of π). The commutative diagram associated with this definition is included below.

$$\begin{array}{ccccc} \mathbf{v} & \xrightarrow{\quad\pi\quad} & \pi(\mathbf{v}) & \xrightarrow{\quad\pi\quad} & \pi^2(\mathbf{v}) = \pi(\mathbf{v}) \\ \downarrow{\scriptstyle \varphi_V} & & \downarrow{\scriptstyle \varphi_V} & & \downarrow{\scriptstyle \varphi_V} \\ \mathbf{x} & \xrightarrow{\quad P\quad} & P\mathbf{x} & \xrightarrow{\quad P\quad} & P^2\mathbf{x} = P\mathbf{x} \end{array}$$

Note that the matrix representation of π is a matrix P satisfying $P^2 = P$; such matrices are known as projection matrices. A simple example of a projection map and its matrix representation in the canonical basis for $\mathbb{E}^3$ is given below.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \overset{\pi}{\mapsto} \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}, \qquad P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Note that in this case also the matrix representation is singular.

17.2 Automorphisms

Definition 17.3 An invertible endomorphism $f\colon V \to V$ is called an **automorphism**.

Automorphisms can also be seen as isomorphisms from a vector space to itself. We denote the set of automorphisms as $\mathrm{Aut}(V)$. The following statements are immediate. We leave the proofs as an exercise.

Proposition 17.4 A linear transformation $f : V \to V$ is an automorphism if and only if $\ker f = \{\mathbf{0}_V\}$.

Corollary 17.5 A linear transformation $f : V \to V$ is an automorphism if and only if f has full rank.

Proposition 17.6 The inverse of an automorphism is an automorphism.

Proposition 17.7 $\mathrm{Aut}(V)$ is a subspace of $\mathcal{L}(V)$.

R Since $\mathrm{Aut}(V)$ satisfies the same properties as $\mathcal{L}(V)$ with respect to the additional operation of composition, it is a non-commutative ring. Since existence of an inverse is satisfied with respect to composition, $(\mathrm{Aut}(V), \circ)$ is a group.

The matrix representations of maps $f \in \mathrm{Aut}(V)$ are invertible, since all f are invertible (see Proposition 16.3). We can use this fact to show that automorphisms are transformations that preserve rank under composition. First, we need the following rank result regarding the composition of two maps.

Proposition 17.8 Let $f : V \to U, g : U \to W$. Then

$$\text{rank } g \circ f \leq \min\{\text{rank } f, \text{rank } g\}.$$

Proof. We have

$$\text{im } g \circ f = g(f(V)) \subseteq g(U) = \text{im } g \Longrightarrow \text{rank } g \circ f \leq \text{rank } g.$$

To obtain the other bound, we define the restriction map $\tilde{g} : \text{im } f \to W$, via $\tilde{g}(\mathbf{u}) = g(\mathbf{u})$ for all $\mathbf{u} \in \text{im } f$. Then $\text{rank } \tilde{g} \leq \min\{\dim \text{im } f, \dim W\}$, by standard rank inequalities. But $\text{im } \tilde{g} = \text{im } g \circ f$ so that

$$\text{rank } g \circ f = \text{rank } \tilde{g} \leq \text{rank } f$$

and the result follows. ■

Proposition 17.9 Let $f \in \mathcal{L}(V, W)$. Let $g_1 \in \mathrm{Aut}(V)$ and $g_2 \in \mathrm{Aut}(W)$. Then

$$\text{rank } g_1 \circ f = \text{rank } f, \qquad \text{rank } f \circ g_2 = \text{rank } f.$$

Proof. Let $r_1 = \text{rank } g_1 \circ f$. Then, but Proposition 17.8,

$$r_1 \leq \min\{\text{rank } g_1, \text{rank } f\} \Longrightarrow r_1 \leq \text{rank } f.$$

Since $g_1 \in \mathrm{Aut}(V)$, g_1 is invertible. Then, using again Proposition 17.8,

$$\text{rank } f = \text{rank } g_1^{-1} \circ (g_1 \circ f) \leq \min\left\{\text{rank } g_1^{-1}, \text{rank } g_1 \circ f\right\} \Longrightarrow \text{rank } f \leq \text{rank } g_1 \circ f \Longrightarrow \text{rank } f \leq r_1.$$

Hence, we have $r_1 = \text{rank } f$. The second statement follows similarly. ■

17.3 Canonical forms

Let us recall the canonical form obtained for the case of linear maps $f : V \to W$ (see Proposition 15.5):

$$A^{\text{can}} = \begin{bmatrix} I_r & O_{r,n-r} \\ O_{m-r,r} & O_{m-r,n-r} \end{bmatrix}.$$

More precisely, any matrix representation A of f is equivalent to the above canonical form:

$$A^{\text{can}} = M^{-1}AN,$$

for some matrices $M \in \mathbb{F}^{m\times m}, N \in \mathbb{F}^{n\times n}$.
This canonical form extends to the case of endomorphisms $f : V \to V$, provided we have the freedom to choose different bases for the domain and codomain. However, in the case of endomorphisms, the domain and codomain are identical, so working with a single basis set is both natural and efficient in applications. This constraint leads to an interesting question: can we find a basis B_V of an n-dimensional vector space V such that the matrix representation is diagonal (possibly including zeros, depending on the rank of the transformation):

$$A^{\text{can}} = \begin{bmatrix} D & O_{r,n-r} \\ O_{n-r,r} & O_{n-r,n-r} \end{bmatrix}, \quad D \in \mathbb{F}^{r\times r}.$$

Should this be possible, we would have any matrix representation of f expressed as

$$A^{\text{can}} = M^{-1}AM,$$

for some matrix $M \in \mathbb{F}^{n\times n}$. This expression describes another equivalence relation on $\mathbb{F}^{n\times n}$, which we include in the following definition.

Definition 17.4 — Matrix similarity. We say matrices $A, B \in \mathbb{F}^{n\times n}$ are similar if there exists an invertible matrix $M \in \mathbb{F}^{n\times n}$ such that

$$B = M^{-1}AM.$$

If we require that B has the above canonical form A^{can}, i.e., B is diagonal, then

$$AM = MB \Longrightarrow A\mathbf{c}_i(M) = b_{ii}\mathbf{c}_i(M) \qquad (i = 1, 2, \dots, n),$$

where $\mathbf{c}_i(M)$ denotes the ith column of M. This is in fact a familiar relation, particularly when written in the form

$$A\mathbf{v}_i = \lambda_i\mathbf{v}_i \qquad (i = 1, 2, \dots, n).$$

The above is referred to as **the eigenvalue problem** for A, with λ_i an eigenvalue and $\mathbf{v}_i$ the corresponding eigenvector. We refer to $(\lambda_i, \mathbf{v}_i)$ as an **eigenpair**, or **eigensolution**, as the above requirement is generally viewed as a (non-linear) equation. This discussion justifies the following definition.

Definition 17.5 A matrix $A \in \mathbb{R}^{n\times n}$ is said to be **diagonalisable** if it is similar to a diagonal matrix.

The similarity requirement implies that there should exist an invertible matrix M such that $B = M^{-1}AM$, whith columns $\mathbf{c}_i(M)$ that need to be linearly independent (for invertibility to hold). This implies that the eigenvectors need to form a basis for $\mathbb{R}^n$ and most clearly cannot be the zero vector. However, there are cases where this is not possible, as the next example shows.

Example 17.3 Consider the eigenproblem

$$A\mathbf{v} := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \lambda \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \iff \begin{cases} v_1 + v_2 & = \lambda v_1 \\ v_2 & = \lambda v_2 \end{cases} \iff \begin{cases} v_1 & = c \\ v_2 & = 0 \end{cases} \iff \mathbf{v} = c\begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

where $c \in \mathbb{R} \setminus \{0\}$ is arbitrary. While it appears that we have infinitely-many solutions, only one of them can be a column of M, as otherwise we would obtain a singular matrix. Hence, there exists a single eigenvector:

$$\mathbf{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

where we made the choice $c = 1$, for simplicity. Therefore, the desired matrix M does not exist, due to the lack of a second linearly independent eigenvector. Hence A is not diagonalisable.

In the next lecture, we will study in detail the eigenvalue problem. As a preliminary result, the above example highlighted the fact that square matrices are not guaranteed to be diagonalisable, i.e., they are not guaranteed to be similar to a diagonal matrix. Hence, diagonal form cannot be viewed as a canonical form, unless we restrict the set of square matrices in a suitable way[1]. A canonical form that replaces the diagonal form is the block-diagonal form known as the **Jordan normal form**. For now, we highlight alternative forms: these can be viewed as 'modern day canonical forms', with square matrices written in factored form, with factors having suitable structures.

17.3.1 Modern canonical forms

We list below a few factorisations (also known as decomposition) that are commonly used in practice, but also as analytical tools. We restrict our attention to real matrices: for complex matrices, please consult the references provided.

The Singular Value Decomposition (SVD)

This is a variation on the canonical form derived for homomorphisms.

Let $m, n \in \mathbb{N}$ with $m > n$. For any matrix $A \in \mathbb{R}^{m\times n}$ there exist orthogonal matrices $U \in \mathbb{R}^{m\times m}, V \in \mathbb{R}^{n\times n}$ such that

$$U^T AV = \Sigma,$$

where $\Sigma \in \mathbb{R}^{m\times n}$ is the diagonal matrix

$$\Sigma = \left[\begin{array}{ccc} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \hline & O_{(m-n)\times n} & \end{array}\right],$$

where $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_n \geq 0$ are called the singular values of A. The index of the smallest non-zero singular value is the rank of A.

R By taking the transpose of A, one can see that a similar decomposition holds for the case $m \leq n$.

The Schur decomposition

Let $A \in \mathbb{R}^{n\times n}$. There exists an orthogonal matrix Q such that

$$Q^T AQ = T = D + U,$$

where D is block diagonal with 1×1 and 2×2 blocks and U is strictly upper triangular.

Remarks:

- A complex conjugate pair of eigenvalues of A corresponds to a 2×2 block of D.
- If A has real eigenvalues, then D is diagonal.
- If A is symmetric, then $U = 0$ and $T = D$ is diagonal.

The Hessenberg decomposition

Let $A \in \mathbb{R}^{n\times n}$. There exists an orthogonal matrix Q such that

$$Q^T AQ = H,$$

where H is a Hessenberg matrix, i.e., an upper triangular matrix with an additional sub-diagonal.

[1] We will see later that symmetric matrices are diagonalisable.

LECTURE 18

Eigenvalues and eigenvectors

18.1 Invariant subspaces

Let $f \in \mathcal{L}(V)$ and let U denote a subspace of V. A natural question to ask is what set f will map U to. Is there any relationship that we can establish between $f(U)$ and U, e.g., $f(U) = U$ or $f(U) \subset U$? The following example shows that, in general, we cannot expect to be able to make a general statement in this respect.

Example 18.1 Let $V = \mathbb{R}^2$ and let

$$f(\mathbf{v}) = A\mathbf{v}, \quad A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Define $U_1 = \text{span}\{\mathbf{e}_1\}, U_2 = \text{span}\{\mathbf{e}_2\}$. Then

$$f(\mathbf{e}_1) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad f(\mathbf{e}_2) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Hence, $f(U_1) = U_1$, but $f(U_2) \neq U_2$; in fact, $f(U_2) \cap U_2 = \{\mathbf{0}\}$.

This example shows that the relationship $f(U) \subseteq U$ is special. This observation justifies the following definition.

Definition 18.1 — Invariant subspace. Let $V(\mathbb{F})$ be a non-trivial vector space and let $f \in \mathcal{L}(V)$. We say a subspace U is invariant under f, or f-invariant, if $f(U) \subseteq U$, i.e., $f(\mathbf{u}) \in U$ for all $\mathbf{u} \in U$.

An immediate consequence of the invariance property is that we can restrict our study of f to its action on invariant subspaces.

Proposition 18.1 Let $f \in \mathcal{L}(V)$ and let $U \leq V$ be an f-invariant subspace. Let $\tilde{f}$ be defined via $\tilde{f}(\mathbf{u}) = f(\mathbf{u})$ for all $\mathbf{u} \in U$. Then $\tilde{f} \in \mathcal{L}(U)$.

The linear map $\tilde{f}$ in the above proposition is called the restriction of f to U and is denoted by $\tilde{f} = f\,|_U$.

Some subspaces are invariant under any $f \in \mathcal{L}(V)$.

Exercise 18.1 Show that $U \leq V$ is invariant under $f \in \mathcal{L}(V)$ for the following choices of U:

1. $U = \{\mathbf{0}\}$;
2. $U = V$;
3. $U = \ker f$;
4. $U = \mathrm{im}\ f$.

We would like to identify other invariant subspaces. Of particular interest will be one-dimensional invariant subspaces.

Motivated by Example 19.2, let us examine the invariance property for the following simple choice of subspace: let $\mathbf{v} \in V \setminus \{\mathbf{0}\}$ and let $U = \mathrm{span}\{\mathbf{v}\}$. Then any non-zero $\mathbf{u} \in U$ has the form $\mathbf{u} = a\mathbf{v}$ for some $a \in \mathbb{F} \setminus \{0\}$. Invariance would then require that $f(\mathbf{u}) \in U$, i.e., $f(a\mathbf{v}) = b\mathbf{v}$, for some $b \in \mathbb{F}$. Using the linearity of f we find

$$f(a\mathbf{v}) = b\mathbf{v} \iff af(\mathbf{v}) = b\mathbf{v} \iff f(\mathbf{v}) = \lambda\mathbf{v}, \qquad \lambda \in \mathbb{F}.$$

Hence, if there exists $\mathbf{v} \in V$ such that the above relation holds, then $U = \mathrm{span}\{\mathbf{v}\}$ is f-invariant. This result justifies the following definition.

Definition 18.2 Let $f \in \mathcal{L}(V)$, where $V(\mathbb{F})$ is a finite-dimensional vector space. Let $\lambda \in \mathbb{F}, \mathbf{v} \in V \setminus \{\mathbf{0}\}$ satisfy the equation

$$f(\mathbf{v}) = \lambda\mathbf{v}. \tag{18.1}$$

Then

- λ is called an **eigenvalue of** f;
- $\mathbf{v}$ is called **an eigenvector of** f;
- the pair $(\lambda, \mathbf{v})$ will be referred to as an **eigenpair of** f or an **eigensolution of** 18.1;
- the set of eigenvalues is called the **spectrum of** f: $\mathrm{sp}f := \{\lambda \in \mathbb{F} : f(\mathbf{v}) = \lambda\mathbf{v}, \mathbf{v} \in V\}$;
- equation 18.1 is called the **eigenvalue equation**.

We emphasise the following facts ensuing from, or stated in, the definition:

- eigenvectors are, by definition, non-zero vectors; without this restriction, any scalar $a \in \mathbb{F}$ would be an eigenvalue, since $f(\mathbf{0}) = \mathbf{0} = a \cdot \mathbf{0}$;
- eigenvectors are unique up to a multiplicative constant: if $\mathbf{v}$ is an eigenvector, then $a\mathbf{v}$ is also an eigenvector, for any $a \in \mathbb{F}$, since $f(a\mathbf{v}) = af(\mathbf{v}) = a(\lambda\mathbf{v}) = \lambda(a\mathbf{v})$;
- an eigenvector is associated with a *single* eigenvalue, since $A\mathbf{v} = \lambda\mathbf{v} = \lambda'\mathbf{v} \implies (\lambda - \lambda')\mathbf{v} = \mathbf{0} \implies \lambda = \lambda'$, given that $\mathbf{v} \neq \mathbf{0}$;
- the eigenvalue equation is nonlinear, since the unknowns λ and $\mathbf{v}$ are multiplied together on the right; this means that we expect $\lambda = \lambda(\mathbf{v})$ and $\mathbf{v} = \mathbf{v}(\lambda)$: this explains the term **eigenpair**.

Before we investigate further this type of f-invariance, we need to establish if the eigenvalue equation for f has any solutions. The answer will depend on the choice of field $\mathbb{F}$.

18.2 The eigenvalue problem: matrix formulations

Let us consider the matrix representation of 18.1. The term on the right-hand side can be written as $\lambda\mathbf{v} = (\lambda id_{\mathrm{V}})(\mathbf{v})$, i.e., a multiple of the identity. Then the matrix representation of 18.1 is

$$A\mathbf{x} = (\lambda I)\mathbf{x} = \lambda\mathbf{x}, \tag{18.2}$$

where $A \in \mathbb{F}^{n\times n}$ is the matrix representation of f and $\mathbf{x} = \varphi_V^{-1}(\mathbf{v}) \in \mathbb{F}^n$, with $\lambda \in \mathbb{F}$. We will study this representation, in order to establish results for the eigenvalue problem 18.1.

Let us re-write 18.2 as follows:

$$A\mathbf{x} = \lambda I\mathbf{x} \iff (\lambda I - A)\mathbf{x} = \mathbf{0}.$$

The above formulation suggests two distinct approaches:

1. Assuming λ is known, one can view this relation as a linear system of equations with coefficient matrix $\lambda I - A$, with zero right hand side. Hence, a non-trivial solution $\mathbf{x} \in \mathbb{F}^n \setminus \{\mathbf{0}\}$ exists if and only if $\det(\lambda I - A) = 0$. This is an eigenvalue-only characterisation! While it seems awkward, it allows us to analyse the existence and properties of eigenvalues independently of the corresponding eigenvectors.
2. Assuming λ is known, one can view eigenvectors as being elements in a certain kernel: $\mathbf{x} \in \ker(\lambda I - A)$. Therefore, the study of such subspaces is useful in providing a description of eigenvectors.

These observations indicate that the study (and computation) of eigenvalues should precede that of eigenvectors. Indeed, this is our approach below.

18.3 Eigenvalues

18.3.1 The characteristic polynomial

The choice of field in our discussion is either $\mathbb{F} = \mathbb{C}$ or $\mathbb{F} = \mathbb{R}$.

Definition 18.3 Let $A \in \mathbb{F}^{n\times n}$. The characteristic polynomial of A is the polynomial $p_A \in \mathcal{P}_n(\mathbb{F})$ defined via

$$p_A(t) = \det(tI - A).$$

We note that, by the properties/definition of the determinant, the polynomial p_A is indeed a polyonimal of degree n, which is also **monic**, i.e., it leading coefficient equal to one; its general form is included below:

$$p(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_{n-1} t^{n-1} + t^n.$$

With this notation in place, the eigenvalue problem (18.2) is equivalent to the following problem:

$$\text{find } \lambda \in \mathbb{F} \text{ such that } p_A(\lambda) = 0.$$

Hence, the eigenvalues of A are the roots (or zeros) of the charateristic polynomial of A.

R If $\mathbb{F} = \mathbb{C}$, then $p_A \in \mathcal{P}_n(\mathbb{C})$, while if $\mathbb{F} = \mathbb{R}$, then $p_A \in \mathcal{P}_n(\mathbb{R})$. In other words, if A is a complex (real) matrix, then the coefficients of the characteristic polynomial p_A are complex (real). This distinction is important, as outlined in the discussion below.

18.3.2 Eigenvalues of complex matrices

Let $A \in \mathbb{C}^{n\times n}$. Then $p_A \in \mathcal{P}_n(\mathbb{C})$. Let us consider the existence of roots of p_A. We recall the following fundamental results for polynomials with complex coefficients.

Theorem 18.2 — Fundamental Theorem of Algebra. Every polynomial $p \in \mathcal{P}_n(\mathbb{C})$ has a zero in $\mathbb{C}$ for any $n \in \mathbb{N}$.

By this result, we can write any polynomial of generic degree n, say $p_n(t)$, in the form

$$p_n(t) = (t - z)p_{n-1}(t),$$

where $z \in \mathbb{C}$ and $p_{n-1}(t)$ is a polynomial of degree $n-1$. Applying the theorem again to $p_{n-1}(t)$, we obtain a further factorisation of $p_n(t)$. This observation leads to the following result.

Proposition 18.3 Let $p \in \mathcal{P}_n(\mathbb{C})$, with $n \in \mathbb{N}$. Then p can be expressed uniquely as the product of n monic linear polynomials $t - z_j$:

$$p(t) = a_n(t - z_1)(t - z_2) \cdots (t - z_n),$$

where $a_n, z_1, \ldots, z_n \in \mathbb{C}$.

This result confirms that the polynomial equation $p(t) = 0$, has n zeros in $\mathbb{C}$. We immediately deduce the following result concerning eigenvalues.

Proposition 18.4 Let $A \in \mathbb{C}^{n \times n}$. Then A has n complex eigenvalues, given by the n roots of $p_A(t)$. In particular, the form of the characteristic polynomial is

$$p_A(t) = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n),$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{C}$.

18.3.3 Eigenvalues of real matrices

Let $A \in \mathbb{R}^{n \times n}$. Then $p_A \in \mathcal{P}_n(\mathbb{R})$. In this case, a similar result to that of Theorem 18.2 does not hold: polynomials with real coefficients do not always have real roots: a simple counter-example is provided by the polynomial $p(t) = t^2 + 1$. In this case, a result that provides a description of the roots of p_A is provided by the factorisation of polynomials with real coefficients. First, we note the following type of quadratic polynomial with real coefficients.

Definition 18.4 A monic quadratic polynomial with real coefficients $p \in \mathcal{P}_2(\mathbb{R})$ is called irreducible if it has the form

$$p(t) = (t - z)(t - \bar{z}) = t - 2\mathrm{Re}z\ t + |z|^2,$$

with $z \in \mathbb{C}$ and $\mathrm{Im}(z) \neq 0$.

R The term *irreducible* refers to the fact that a monic polynomial $p \in \mathcal{P}_2(\mathbb{R})$ can only be factorised as the product of two monic polynomials in $\mathcal{P}_1(\mathbb{C})$ and not in $\mathcal{P}_1(\mathbb{R})$.

The main result regarding factorisation of polynomials over $\mathbb{R}$ is included below.

Proposition 18.5 Let $p \in \mathcal{P}_n(\mathbb{R})$, with $n \in \mathbb{N}$. Then p can be expressed uniquely as the product of monic linear and irreducible quadratic polynomials with real coefficients in the form

$$p(t) = a_n(t - x_1)(t - x_2) \cdots (t - x_m)(t^2 + b_1 x + c_1)(t^2 + b_2 x + c_2) \cdots (t^2 + b_k x + c_k),$$

where $a_n, x_i, b_j, c_j \in \mathbb{R}$, $i = 1, \ldots, m$ and $j = 1, \ldots, k$, with $n = m + 2k$.

Hence, a polynomial of degree n with real coefficients has
- $0 \leq m \leq n$ real roots;
- k pairs of complex-conjugate roots, with $0 \leq k \leq \frac{n-m}{2}$.

The above representation allows us to provide a description (albeit theoretical) of the real eigenvalues of a real matrix A.

Proposition 18.6 Let $A \in \mathbb{R}^{n \times n}$. Then A has a number $0 \leq m \leq n$ of real eigenvalues equal to the number of linear factors in the factorisation of its characteristic polynomial p_A.

R Note that the case real matrices $A \in \mathbb{R}^{n\times n}$ corresponds to endomorphisms f defined on real vector spaces. Since we are interested in the invariance of one-dimensional subspaces of a real vector space $V(\mathbb{R})$, we cannot take into account the complex roots of the characteristic polynomial of A, as the scalars arising in the invariance property (i.e., the eigenvalues) need to be real.

We summarise the above discussion as follows:
- We have at most n roots of p_A, hence at most n eigenvalues, possibly non-distinct.
- If $\mathbb{F} = \mathbb{C}$, then we will always have n eigenvalues in $\mathbb{C}$, some possibly real.
- If $\mathbb{F} = \mathbb{R}$, then we are not guaranteed to have n eigenvalues in $\mathbb{R}$.

This means that if we pose the eigenvalue problem over $\mathbb{C}$, then we are guaranteed to have n eigenvalues in $\mathbb{C}$, possibly with some in $\mathbb{R}$. For this reason, we reformulate the eigenvalue problem for $f \in \mathcal{L}(V)$ as follows:

$$\text{find } \lambda \in \mathbb{C} \text{ and } \mathbf{v} \in \mathbb{C}^n \text{ such that } f(\mathbf{v}) = \lambda\mathbf{v}.$$

We note that in the above formulation we necessarily have $V = V(\mathbb{C})$.

This formulation leads to the obvious question: how do we deal with the case of real linear maps? By the above results, we could have real maps that yield characteristic polynomials with
- no real roots (if n is even, this is a possibility); in this case, f has no eigensolutions over $\mathbb{R}$ and hence no invariant subspaces of dimension one;
- n real roots: in this case, we could pose the problem over $\mathbb{F} = \mathbb{R}$: this would be natural and also efficient, in practice. We know that such maps exist: an example is given by the identity map, which has n real eigenvalues, all equal to one. Later, we will identify other real linear maps with n real eigenvalues.

One way of reconciling the real case with the above complex formulation is to note that any real map induces a complex map: $f \leftrightarrow f + i\cdot o$, where o is the zero map over $\mathbb{C}$. We can then seek eigensolutions over $\mathbb{C}$: if all the eigenvalues turn out to be real, we can also choose the eigenvectors to be real and reformulate the eigenvalue problem over $\mathbb{R}$, *a posteriori*.

18.3.4 Algebraic multiplicity

If the characteristic polynomial has repeated roots, then its general form over $\mathbb{C}$ is given by

$$p(t) = a_n(t-\lambda_1)^{\alpha_1}(t-\lambda_2)^{\alpha_2}\cdots(t-\lambda_\ell)^{\alpha_\ell},$$

where $\alpha_1 + \alpha_2 + \cdots + \alpha_\ell = n$. Let us define the map $\alpha : \mathrm{sp}f \mapsto \{1, 2, \ldots, n\}$ via $\alpha(\lambda_k) = \alpha_k$.

Definition 18.5 Let V be an n-dimensional vector space and let $f \in \mathcal{L}(V)$. The algebraic multiplicity of an eigenvalue λ_k of f is $\alpha(\lambda_k) = \alpha_k$.

Note that we can indeed have $\alpha(\lambda) = n$: this is the case of the eigenvalues of the identity map, which are all equal (to one); in fact, any map $f = c \cdot id$ for some $c \in \mathbb{F}$ will also satisfy this property. Note also that if $\alpha(\lambda) = 1$ for all $\lambda \in \mathrm{sp}f$, then we have n distinct eigenvalues and therefore n distinct eigenvectors. If $\alpha(\lambda) > 1$ for some λ, i.e., at least one eigenvalue is a repeated eigenvalue, then there may be situations where the total number of eigenvectors is less than n: see Example 19.2. This is an important distinction which we will discuss in the next lecture.

LECTURE 19

Eigenvalues and eigenvectors (cont.)

19.1 Eigenvectors

We consider again the eigenvalue equation

$$f(\mathbf{v}) = \lambda \mathbf{v},$$

where $f \in \mathcal{L}(V)$, $\lambda \in \mathbb{F}$. Let us recall also its matrix formulation in the two equivalent forms considered previously:

$$A\mathbf{x} = \lambda \mathbf{x} \iff (\lambda I - A)\mathbf{x} = \mathbf{0}.$$

19.1.1 Eigenspaces

The first observation is that $\mathbf{v}$ belongs to a certain subspace of V.

Proposition 19.1 Let $(\lambda, \mathbf{v})$ be an eigenpair of f. Then $\mathbf{v} \in E_\lambda := \ker(f - \lambda id_{\mathrm{V}})$. Moreover, $E_\lambda \leq V$.

Proof. Since $\lambda \mathbf{v} = (\lambda id_{\mathrm{V}})\mathbf{v}$, we find that

$$f(\mathbf{v}) - (\lambda id_{\mathrm{V}})(\mathbf{v}) = \mathbf{0} \iff (f - \lambda id_{\mathrm{V}})(\mathbf{v}) = \mathbf{0} \iff \mathbf{v} \in \ker(f - \lambda id_{\mathrm{V}}).$$

Finally, since$f, id_{\mathrm{V}} \in \mathcal{L}(V)$, the map $f - \lambda id_{\mathrm{V}} \in V$, by closure in $\mathcal{L}(V)$. The result then follows, as the kernel of an endomorphism on V is a subspace of V. ■

The subspace property in the previous result suggests the next definition.

Definition 19.1 — Eigenspace. The subspace E_λ is the eigenspace of f associated with eigenvalue λ.

Note that for any λ, E_λ is non-trivial, since it contains at least one non-zero vector: an eigenvector associated with λ. This means that dim $E_\lambda \geq 1$. Let us derive further properties of eigenspaces.

Proposition 19.2 Let $(\lambda, \mathbf{v}), (\lambda', \mathbf{v}')$ denote two distinct eigenpairs. Then $E_\lambda \cap E_{\lambda'} = \{\mathbf{0}_{\mathrm{V}}\}$.

Proof. Assume, for a contradiction, that there exists a nonzero $\mathbf{u}$ such that $\mathbf{u} \in E_\lambda \cap E_{\lambda'}$. Then

$$f(\mathbf{u}) = \lambda \mathbf{u} = \lambda' \mathbf{u} \Longrightarrow (\lambda - \lambda')\mathbf{u} = \mathbf{0}_{\mathrm{V}} \Longrightarrow \lambda = \lambda',$$

which is the contradiction we sought. ■

We immediately obtain the following corollary.

Corollary 19.3 Eigenvectors corresponding to different eigenvalues (i.e., from different eigenspaces) are linearly independent.

Another corollary is included below.

Corollary 19.4 Let dim $E_\lambda = 1$ for all $\lambda \in \mathrm{sp} f$. Then V is a direct sum of eigenspaces:

$$V = \bigoplus_{\lambda \in \mathrm{sp} f} E_\lambda.$$

In particular, the eigenvectors form a basis for V.

Can we actually have dim $E_\lambda > 1$? The answer is provided by the following example:

Example 19.1 Let $V(\mathbb{F})$ be an n-dimensional vector space and let $f = id_{\mathrm{V}}$. Then its matrix representation is I_n, which has a single eigenvalue $\lambda = 1$, with algebraic multiplicity n. Moreover, each canonical vector $\mathbf{e}_i \in \mathbb{R}^n$ is an eigenvector for λ, so that $E_\lambda = \mathbb{R}^n$ and dim $E_\lambda = n$.

It is clear from this example that the algebraic multiplicity of λ is related to the dimension of E_λ. Let us look at this more closely.

19.1.2 Geometric multiplicity

Definition 19.2 The **geometric multiplicity** of λ is denoted by $\gamma(\lambda)$ and is defined to be the dimension of its associated eigenspace: $\gamma(\lambda) := \dim \ker E_\lambda$.

Let $B_\lambda = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ be a basis for E_λ. Then each element $\mathbf{v}_j$ of B_λ is an eigenvector of f. Therefore, the geometric multiplicity of λ can be viewed as the number of linearly independent eigenvectors associated with λ.

The following result provides some initial insight into the existence of eigenspaces of dimension greater than one.

Proposition 19.5 Let $f \in \mathcal{L}(V(\mathbb{F}))$, where $V(\mathbb{F})$ is an n-dimensional vector space. Then every eigenvalue λ of f has geometric multiplicity no greater than the algebraic multiplicity: $\gamma(\lambda) \leq \alpha(\lambda)$.

Proof. Let $1 \leq r = \gamma(\lambda)$ and $B_\lambda = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ be a basis for E_λ, i.e., $f(\mathbf{v}_k) = \lambda \mathbf{v}_k$ for $k = 1, 2, \ldots, r$. Let us choose now a basis B for V containing B_λ. Then the matrix representation of f takes the form

$$A = \begin{bmatrix} \lambda I_r & B \\ O & C \end{bmatrix}.$$

Hence, using the properties of determinants,

$$p_A(t) = \det(tI - A) = \det(tI_r - \lambda I_r) \cdot \det(tI_{n-r} - C) = (t - \lambda)^r p_C(t),$$

where $p_C \in \mathcal{P}_{n-r}(\mathbb{F})$ is the characteristic polynomial of C. Therefore, $\alpha(\lambda) \geq r$, since there are at least r factors $t - \lambda$ of $p_A(t)$. Thus, $\gamma(\lambda) = |B_\lambda| = r \leq \alpha(\lambda)$. ∎

An important consequence of the above result is that the sum of geometric multiplicities is no greater than n:

$$\sum_{k=1}^{n} \gamma(\lambda_k) \leq \sum_{k=1}^{n} \alpha(\lambda_k) = n.$$

Since the geometric multiplicity $\gamma(\lambda)$ can be viewed as the number of linearly independent eigenvectors associated with λ, we deduce that the total number of linearly independent eigenvectors of a linear map can be less than n.

Can we have indeed $\gamma(\lambda) \le \alpha(\lambda)$ for some λ? The answer is yes; here is an example we encountered when we discussed invariant subspaces and which we recall below.

Example 19.2 Let $V = \mathbb{R}^2$ and let

$$f(\mathbf{v}) = A\mathbf{v}, \quad A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Then the eigenvalues are $\lambda_1 = \lambda_2 = \lambda = 1$, but we find that $\gamma(\lambda) = 1 < \alpha(\lambda) = 2$, since

$$(\lambda I - A)\mathbf{v} = \mathbf{0} \iff \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix} \mathbf{v} = \mathbf{0} \iff \begin{cases} v_1 \in \mathbb{R} \\ v_2 = 0 \end{cases} \iff \mathbf{v} \in \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\} =: E_\lambda.$$

Note. This is an example of so-called **Jordan block** (of size 2, with eigenvalue $\lambda = 1$).

There are plenty examples where the geometric multiplicity can take any value between 1 and n.

Exercise 19.1 The following matrices have a single eigenvalue $\lambda = 1$. In each case, find the geometric multiplicity $\gamma(\lambda)$.

$$\text{i. } A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \text{ii. } B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

19.2 Canonical forms

In Lecture 17 we introduced the concepts of similarity and diagonalisation. Recall that a square matrix A is said to be diagonalisable if it is similar to a diagonal matrix D:

$$A = M^{-1}DM \iff D = MAM^{-1},$$

for some invertible M. In this case, D represents the canonical form of the equivalence class of matrices similar to A. We say that the associated endomorphisms are diagonalisable. The following result provides a characterisation of diagonalisation of endomorphisms $f \in \mathcal{L}(V(\mathbb{F}))$, and therefore equivalently of square matrices $A \in \mathbb{F}^{n\times n}$.

Proposition 19.6 Let $V(\mathbb{F})$ be an n-dimensional vector space with basis $B_V = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. An endomorphism $f : V \to V$ has a diagonal matrix representation relative to B_V if and only $\mathbf{v}_j$ is an eigenvector for f for all $j = 1, \ldots, n$.

Proof. $\boxed{\Longrightarrow}$ Assume that f has a diagonal matrix representation $D \in \mathbb{F}^{n\times n}$ with diagonal entries λ_i. Equivalently, we have $[D]_{ij} = \lambda_i \delta_{ij}$ for $i, j = 1, \ldots, n$. Consider

$$f(\mathbf{v}_j) = \sum_{i=1}^{n} y_i \mathbf{v}_i.$$

By Proposition 13.1, $\mathbf{y} = D\mathbf{x}$, where $\mathbf{x} = \mathbf{e}_j$, given that we are only evaluating the jth vector in the basis B_V. We find

$$\mathbf{y} = D\mathbf{e}_j = \mathbf{c}_j(D) = \lambda_j \mathbf{e}_j \Longrightarrow y_i = \lambda_j \delta_{ij}.$$

Therefore,

$$f(\mathbf{v}_j) = \sum_{i=1}^{n} y_i \mathbf{v}_i = \sum_{i=1}^{n} \lambda_j \delta_{ij} \mathbf{v}_i = \lambda_j \mathbf{v}_j,$$

which indicates that $\mathbf{v}_j$ is an eigenvector of f, with corresponding eigenvalue λ_j.
$\boxed{\Longleftarrow}$ Assume now that A is the matrix representation relative to a basis of eigenvectors B_V. Noting that $f(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$, by definition of matrix representation, we get

$$f(\mathbf{v}_i) = \sum_{j=1}^{n} a_{ji} \mathbf{v}_j.$$

On the other hand,

$$f(\mathbf{v}_i) = \lambda_i \mathbf{v}_i = \sum_{j=1}^{n} \lambda_i \delta_{ji} \mathbf{v}_j.$$

By uniqueness of representation of $f(\mathbf{v}_i)$ in the basis B_V, we find $a_{ji} = \lambda_i \delta_{ji}$ for all j, which implies that A is a diagonal matrix with the eigenvalues as diagonal entries. ■

R Note that f does not have to be invertible in order for its matrix representation to have a diagonal canonical form.

This result implies that the only situation where we have a canonical form for an endomorphism defined on an n-dimensional vector spaces corresponds to the case where we have n eigenvectors. In turn, this implies that we require

$$\sum_{k=1}^{n} \gamma(\lambda_k) = n.$$

This observation justifies the following terminology.

Definition 19.3 — Defective matrix. We say a matrix $A \in \mathbb{F}^{n \times n}$ is defective if it does not have n linearly independent eigenvectors.

It is important to realise that defective matrices arise 'naturally' in applications and the lack of a complete set of eigenvectors cannot be avoided by a change of basis. In fact, using this very concept, we can see that the matrix in Example 19.2 can be viewed as a canonical form for an equivalence class of 2×2 matrices which have a single eigenvector associated with the repeated eigenvalue $\lambda = 1$.

The following results are obvious corollaries.

Corollary 19.7 Defective matrices are not diagonalisable.

Corollary 19.8 If the eigenvalues of an endomorphism f are distinct, then f is diagonalisable over $\mathbb{C}$.

Corollary 19.9 If the eigenvalues of an endomoprhism defined on an n-dimensional vector space have geometric multiplicities summing up to n, then the map is diagonalisable over $\mathbb{C}$.

R Note that for general real matrices we are not guaranteed to have a diagonal canonical form over $\mathbb{R}$ (i.e., a diagonal matrix with real entries), even if the eigenvalues are distinct. This is due to the fact that the characteristic polynomial of a general real matrix cannot be factorised over $\mathbb{R}$, i.e., it cannot be written as a product of real linear polynomials only (see Proposition 18.5).

19.3 Invariant subspaces

We started our discussion of eigenvalues and eigenvectors motivated by the concept of subspace invariance for the case of one-dimensional subspaces. It is therefore fitting to round up our discussion by revisiting this concept and considering the case of subspaces of dimension greater than one.

What is the consequence of invariance of a k-dimensional subspace with regard to matrix representations? To answer this, let us consider the simple case of a two-dimensional f-invariant subspace $U = \text{span}\{\mathbf{u}_1, \mathbf{u}_2\}$ of an n-dimensional vector space V. Let $B = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_3, \dots, \mathbf{v}_n\}$ be a basis for V. Then

$$\begin{cases} f(\mathbf{u}_1) & = a_{11}\mathbf{u}_1 + a_{21}\mathbf{u}_2 \\ f(\mathbf{u}_2) & = a_{12}\mathbf{u}_1 + a_{22}\mathbf{u}_2 \\ f(\mathbf{v}_3) & = b_{13}\mathbf{u}_1 + b_{23}\mathbf{u}_2 + b_{33}\mathbf{v}_3 + \cdots + b_{n3}\mathbf{v}_n \\ & \dots \\ f(\mathbf{v}_n) & = b_{1n}\mathbf{u}_1 + b_{2n}\mathbf{u}_2 + b_{3n}\mathbf{v}_3 + \cdots + b_{nn}\mathbf{v}_n, \end{cases}$$

so that the matrix representation in the basis B is

$$A = \begin{bmatrix} a_{11} & a_{12} & b_{13} & \cdots & \cdots & b_{1n} \\ a_{21} & a_{22} & b_{23} & \cdots & \cdots & b_{2n} \\ 0 & 0 & b_{33} & \cdots & \cdots & b_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & b_{n3} & \cdots & \cdots & b_{nn} \end{bmatrix} = \begin{bmatrix} A_U & B \\ O & C \end{bmatrix},$$

where A_U is the matrix representation of $\tilde{f} = f\,|_U$ and O is the $(n-2) \times 2$ zero matrix. Thus, the matrix representation of a linear map with a two-dimensional invariant subspace has a block upper-triangular form. This is just a special case of a more general result on invariance which we state below.

Proposition 19.10 Let $f \in \mathcal{L}(V)$ and let $U \leq V$, with $\dim U = k, \dim V = n$. Then U is f-invariant if and only if it has a block triangular matrix representation of the form

$$A = \begin{bmatrix} A_U & B \\ O & C \end{bmatrix},$$

where $A_U \in \mathbb{F}^{k\times k}, O \in \mathbb{F}^{(n-k)\times k}, B \in \mathbb{F}^{k\times(n-k)}, C \in \mathbb{F}^{(n-k)\times(n-k)}$.

The following exercise is a corollary of the above result.

Exercise 19.2 Let $f \in \mathcal{L}(V)$ and assume $V = U \oplus W$, with U, W f-invariant subspaces of V. Show that f has a block-diagonal matrix representation:

$$A = \begin{bmatrix} A_U & O \\ O & A_W \end{bmatrix},$$

where A_U, A_W are square matrices of dimensions summing up to $\dim V$.

R Block diagonal matrices are commonly written as **matrix direct sums** $A = A_U \oplus A_W$.

We can use Proposition 19.10 to construct a canonical form for an endomorphism $f \in \mathcal{L}(V)$ over $\mathbb{R}$. Let us assume first that f has a non-defective matrix representation A. For ease of argument, let us assume that the eigenvalues are distinct. By Proposition 18.5, we have in general m real eigenvalues, with $k = (n-m)/2$ complex-conjugate roots. We first need the following result.

Lemma 19.11 Let $(\lambda, \mathbf{v})$ be a complex eigenpair of A, where

$$\lambda = a + ib, \qquad \mathbf{v} = \mathbf{u}_1 + i\mathbf{u}_2,$$

with $a, b \in \mathbb{R}$ with $b \neq 0$ and $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^n \setminus \{\mathbf{0}\}$. Then $\{\mathbf{u}_1, \mathbf{v}\}$ is a linearly independent set in $\mathbb{R}^n$. Moreover,

$$A[\mathbf{u}_1 \ \mathbf{u}_2] = [\mathbf{u}_1 \ \mathbf{u}_2] \begin{bmatrix} a & b \\ -b & a \end{bmatrix}.$$

Proof. First, note that if $\{\mathbf{u}_1, \mathbf{v}\}$ is a linearly dependent set, then $\mathbf{u}_1 = c\mathbf{u}_2$, for some $c \in \mathbb{R}$; then $\mathbf{v} = (1 + ic)\mathbf{u}_1$ is an eigenvector and therefore $\tilde{\mathbf{v}} = \mathbf{u}_1 \in \mathbb{R}^n$ is also one. This results in $A\mathbf{u}_1 = (a + ib)\mathbf{u}_1$; comparing imaginary parts, we obtain $b = 0$, a contradiction. Hence, $\{\mathbf{u}_1, \mathbf{v}\}$ is a linearly independent set in $\mathbb{R}^n$.

To derive the second statement, we note that

$$A\mathbf{v} = \lambda\mathbf{v} \Longleftrightarrow A(\mathbf{u}_1 + i\mathbf{u}_2) = (a + ib)(\mathbf{u}_1 + i\mathbf{u}_2) \Longleftrightarrow \begin{cases} A\mathbf{u}_1 = & a\mathbf{u}_1 - b\mathbf{u}_2, \\ A\mathbf{u}_2 = & a\mathbf{u}_2 + b\mathbf{u}_1, \end{cases}$$

after identifying real and imaginary parts. The result follows by writting these identities in matrix form. ∎

Corollary 19.12 Let $A \in \mathbb{R}^{n \times n}$. Let $(\lambda, \mathbf{v})$ be a complex eigenpair of A, where

$$\lambda = a + ib, \qquad \mathbf{v} = \mathbf{u}_1 + i\mathbf{u}_2,$$

with $a, b \in \mathbb{R}$, $b \neq 0$ and $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^n \setminus \{\mathbf{0}\}$. Then $U := \operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2\}$ is an A-invariant subspace of $\mathbb{R}^n$.

Proof. Let $B = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n\}$ be a basis for R^n. Let $\mathbf{u} \in U$. Then

$$A\mathbf{u} = A(\alpha\mathbf{u}_1 + \beta\mathbf{u}_2) = \alpha A\mathbf{u}_1 + \beta A\mathbf{u}_2 = \alpha(a\mathbf{u}_1 - b\mathbf{u}_2) + \beta(a\mathbf{u}_2 + b\mathbf{u}_1) = (a\alpha + b\beta)\mathbf{u}_1 + (a\beta - b\alpha)\mathbf{u}_2 \in U.$$

∎

The results of the lemma and the corollary lead to the following important observation: given a real matrix A, the two complex one-dimensional A-invariant subspaces of $\mathbb{C}^n$ associated with a complex eigenvalue and its complex-conjugate induce a real two-dimensional A-invariant subspace of $\mathbb{R}^n$. This means that we can replace the complex diagonal canonical form of A with a real canonical form that involves a block diagonal matrix, with real blocks of size 1×1, if the corresponding eigenvalue is real, or of size 2×2 for corresponding complex eigenvalues $\lambda, \bar{\lambda}$.

Proposition 19.13 Let $A \in \mathbb{R}^{n \times n}$ be a non-defective matrix with
- m real eigenvalues: $\lambda_1, \ldots, \lambda_m$, $i = 1, \ldots, m$;
- $k = (n - m)/2$ complex-conjugate eigenvalues: $\lambda_j = a_j \pm ib_j$, $j = 1, \ldots, k$.

Then $A = MDM^{-1}$, with $D, M \in \mathbb{R}^{n \times n}$ and with D a block-diagonal matrix with
- m blocks of size 1×1: $\lambda_1, \ldots, \lambda_m$;
- k blocks of size 2×2 of the form

$$\begin{bmatrix} a_j & b_j \\ -b_j & a_j \end{bmatrix}, \qquad j = 1, \ldots, k.$$

The block-diagonal matrix D is the real canonical form of a square real matrix A.

We briefly summarise our findings below: any non-defective matrix can be diagonalised as follows:
- $A \in \mathbb{C}^{n \times n}$ is similar to a complex diagonal matrix, with eigenvalues on the diagonal;
- $A \in \mathbb{R}^{n \times n}$ is similar to a block diagonal matrix, with diagonal entries as described in Prop. 19.13.

LECTURE 20

Symmetry

We finish our discussion of linear transformations by considering the case where the domain and codomain are inner product spaces. This context provides a rich source of additional properties that we can identify for our maps, with a wide range of applications.

In the following, we will assume that our vector spaces are real and finite-dimensional and are equipped with

- inner-products, generically denoted by $\langle \cdot, \cdot \rangle_V$;
- bases which are orthonormal with respect to $\langle \cdot, \cdot \rangle_V$.

We will also denote the standard Euclidean inner product on $\mathbb{R}^n$ by $\langle \cdot, \cdot \rangle_n$, or occasionally by $\langle \cdot, \cdot \rangle_{R^n}$.

20.1 Adjoint maps

Let V, W be inner-product spaces and let $f \in \mathcal{L}(V, W)$. Let us consider defining another linear map $g : W \to V$; to narrow down this task, consider the following two results. The first is essentially Proposition 8.4, using notation relevant to the current topic.

Proposition 20.1 Let $(V, \langle \cdot, \cdot \rangle_V)$ be an inner product space. Let $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ denote a basis for V which is orthonormal with respect to $\langle \cdot, \cdot \rangle_V$. Then, for any $\mathbf{u}, \mathbf{v} \in V$,

$$\langle \mathbf{u}, \mathbf{v} \rangle_V = \langle \mathbf{x}, \mathbf{y} \rangle_n ,$$

where $\mathbf{u} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$ and $\mathbf{v} = y_1\mathbf{v}_1 + y_2\mathbf{v}_2 + \cdots + y_n\mathbf{v}_n$.

Proof. We have, using the linearity of the inner product and the orthonormality of $\mathbf{v}_i$,

$$\langle \mathbf{u}, \mathbf{v} \rangle_V = \langle x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n, y_1\mathbf{v}_1 + \cdots + y_n\mathbf{v}_n \rangle = x_1y_1 + \cdots + x_ny_n = \langle \mathbf{x}, \mathbf{y} \rangle_n .$$

■

The inner product on $\mathbb{R}^n$ can also be given the following convenient form involving the product of two 'matrices': the $1 \times n$ matrix (row vector) $\mathbf{x}^T \in \mathbb{R}^{1\times n}$ and the $n \times 1$ matrix (column vector) $\mathbf{y} \in \mathbb{R}^{n\times 1}$:

$$\langle \mathbf{x}, \mathbf{y} \rangle_n = \mathbf{x}^T\mathbf{y} = \mathbf{y}^T\mathbf{x}.$$

We used the *transpose* notation to denote the row vector, which is seen here as the transpose of the column vector $\mathbf{x}$. Generally, one can define the transpose A^T of a matrix as the matrix with entries $[A^T]_{ij} = [A]_{ji}$.

The following result is a corollary.

Proposition 20.2 Let $f \in \mathcal{L}(V, W)$, where V, W are inner product spaces equipped with orthonormal bases. Let $A \in \mathbb{R}^{m \times n}$ be the matrix representation of f with respect to these bases. Then

$$\langle f(\mathbf{v}), \mathbf{w} \rangle_W = \langle A\mathbf{x}, \mathbf{y} \rangle_m = \mathbf{y}^T A\mathbf{x},$$

where $\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$ and $\mathbf{w} = y_1\mathbf{w}_1 + y_2\mathbf{w}_2 + \cdots + y_m\mathbf{w}_m$.

Proof. The result follows by applying Proposition 20.1 to the evaluation of the inner product $\langle \mathbf{u}, \mathbf{w} \rangle$ where $\mathbf{u} = f(\mathbf{v})$, with $\varphi_V(\mathbf{u}) = A\mathbf{x}$. ■

The expression for the inner product in Proposition 20.2 can be written also in the form

$$\mathbf{y}^T A\mathbf{x} = \mathbf{y}^T (A\mathbf{x}) = (A\mathbf{x})^T \mathbf{y} = \mathbf{x}^T A^T \mathbf{y}.$$

This suggests that we could consider another linear map with matrix representation $A^T \in \mathbb{R}^{n \times m}$. This is described in the following definition.

Definition 20.1 — Adjoint. Let $f \in \mathcal{L}(V, W)$, where V, W are inner product spaces. The adjoint of f is a linear map $f^* : W \to V$ defined via

$$\langle f(\mathbf{v}), \mathbf{w} \rangle_W = \langle \mathbf{v}, f^*(\mathbf{w}) \rangle_V .$$

The concept of adjoint map is well-defined, since it can readily be seen to be the unique map satisfying the relation given by the definition. By the above discussion, the following result holds.

Proposition 20.3 Let $f \in \mathcal{L}(V, W)$, where V, W are inner product spaces equipped with orthonormal bases. Let A be the matrix representation of f. Then the matrix representation of f^* is A^T.

We can actually derive an explicit expression for the action of the adjoint map on a vector $\mathbf{w} \in W$.

Proposition 20.4 Let $f \in \mathcal{L}(V, W)$, where V, W are inner product spaces equipped with orthonormal bases. Then the adjoint of f is given by

$$f^*(\mathbf{w}) = \sum_{i=1}^{n} \langle \mathbf{w}, f(\mathbf{v}_i) \rangle_W \mathbf{v}_i.$$

Proof. Recall first that orthonormal bases allow for a Fourier representation of a vector, with the coefficients written as inner products (see Lecture 8):

$$\mathbf{v} = \sum_{i=1}^{n} \langle \mathbf{v}, \mathbf{v}_i \rangle_V \mathbf{v}_i.$$

Setting $\mathbf{v} = f^*(\mathbf{w})$, we find the coefficients take the form

$$\langle \mathbf{v}, \mathbf{v}_i \rangle_V = \langle f^*(\mathbf{w}), \mathbf{v}_i \rangle_V = \langle \mathbf{w}, f(\mathbf{v}_i) \rangle_W ,$$

where we used the definition of adjoint. ■

20.2 Self-adjoint maps

There are many interesting results one can establish on adjoints; however, in this lecture we will restrict our attention to the case of endomorphisms $f \in \mathcal{L}(V)$. In particular, we wish to study the following concept.

Definition 20.2 — Self-adjoint map. An endomorphism f defined on a finite-dimensional inner product space V is called self-adjoint if $f = f^*$ with respect to the inner product on V.

Note that, by definition, there holds

$$\langle f(\mathbf{v}), \mathbf{w}\rangle_V = \langle \mathbf{v}, f(\mathbf{w})\rangle_V .$$

This expression yields the following characterisation of a self-adjoint map.

Theorem 20.5 Let $f \in \mathcal{L}(V)$ where V is an inner product equipped with an orthonormal basis B. Then f is self-adjoint if and only if its matrix representation relative to the basis B is a symmetric matrix.

Proof. $\boxed{\Longrightarrow}$ Let f be self-adjoint and let A denote its matrix representation relative to the orthonormal basis B. We write the self-adjointness property as

$$\langle f(\mathbf{v}), \mathbf{w}\rangle_V = \langle \mathbf{v}, \mathbf{f}(\mathbf{w})\rangle_V = \langle f(\mathbf{w}), \mathbf{v}\rangle_V .$$

Let $\varphi_V(\mathbf{v}) = \mathbf{x}, \varphi_V(\mathbf{w}) = \mathbf{y}$. By Proposition 20.2,

$$\langle f(\mathbf{v}), \mathbf{w}\rangle_V = \mathbf{y}^T A\mathbf{x}, \quad \langle f(\mathbf{w}), \mathbf{v}\rangle_V = \mathbf{x}^T A\mathbf{y}.$$

On the other hand, since $\mathbf{x}^T A\mathbf{y}$ is a scalar and the transpose of a scalar is the scalar, the above relation can be written as

$$\mathbf{y}^T A\mathbf{x} = \mathbf{x}^T A\mathbf{y} = (\mathbf{x}^T A\mathbf{y})^T = \mathbf{y}^T A^T\mathbf{x} \Longrightarrow A = A^T,$$

and the matrix is symmetric.

$\boxed{\Longleftarrow}$ Let the matrix representation A of f relative to B be symmetric. By Proposition 20.3, the matrix representation of f^* is $A^T = A$. Since the matrix representations are equal, $f = f^*$. ■

Example 20.1 Let $V = \{p \in \mathcal{P}_2(\mathbb{R}) : p(-a) = p(a) = 0\}$ and let $f \in \mathcal{L}(V)$ be the map $f(p) = p + p''$. Let $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ be defined via

$$\langle p, q\rangle = \int_{-a}^{a} p(x)q(x)\mathrm{d}x.$$

Let us find f^*. Let $p, q \in V$. Then

$$\langle p, f^*(q)\rangle = \langle f(p), q\rangle = \int_{-a}^{a} (p(x)+p''(x))q(x)\mathrm{d}x = \int_{-a}^{a} p(x)q(x)\mathrm{d}x + \overbrace{\left[p'(x)q(x)\right]_{-a}^{a}}^{=0} - \int_{-a}^{a} p'(x)q'(x)\mathrm{d}x =$$

$$\int_{-a}^{a} p(x)q(x)\mathrm{d}x - \overbrace{\left[p(x)q'(x)\right]_{-a}^{a}}^{=0} + \int_{-a}^{a} p(x)q''(x)\mathrm{d}x = \int_{-a}^{a} p(x)\left(q(x) + q''(x)\right)\mathrm{d}x,$$

so that $f^*(q) = q + q'' = f(q)$. Hence $f : V \to V$ is self-adjoint with respect to $\langle\cdot,\cdot\rangle$.

20.3 Spectral properties

Let us turn our attention to the eigenvalues of a self-adjoint map. We will continue to discuss the case of real inner product spaces. This is of particular interest, as we learned in the previous lecture that diagonalising maps over $\mathbb{R}$ is not possible. However, this changes if we restrict the set of maps to self-adjoint maps. We start with the following preliminary result.

Lemma 20.6 Let V be a real inner product space and let $f \in \mathcal{L}(V)$ be self-adjoint. Then the following spectral properties hold:

i. the eigenvalues of f are real;
ii. the eigenvectors of f can be chosen to be real.

Proof. i. Let V be a real inner product space and let $f \in \mathcal{L}(V)$ be self-adjoint. Let $A \in \mathbb{R}^{n\times n}$ be the matrix representation of f in some orthonormal basis of V. Then A is symmetric. Let us consider the eigenvalue problem for A. Since in general we are not guaranteed to have eigensolutions in $\mathbb{R}$, let us consider the problem over $\mathbb{C}$: find $\lambda \in \mathbb{C}, \mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$, such that

$$A\mathbf{x} = \lambda\mathbf{x}.$$

First note that λ can be given the following representation:

$$\overline{\mathbf{x}}^T A\mathbf{x} = \lambda\overline{\mathbf{x}}^T\mathbf{x} \iff \lambda = \frac{\overline{\mathbf{x}}^T A\mathbf{x}}{\overline{\mathbf{x}}^T\mathbf{x}} =: \frac{z_1}{z_2},$$

where $\overline{\mathbf{x}}$ denotes taking the complex-conjugate entrywise. Now,

$$z_1 := \overline{\mathbf{x}}^T A\mathbf{x} = \overline{\mathbf{x}}^T(A\mathbf{x}) = \langle\overline{\mathbf{x}}, A\mathbf{x}\rangle = (A\mathbf{x})^T\overline{\mathbf{x}} = \mathbf{x}^T A^T\overline{\mathbf{x}} = \mathbf{x}^T A\overline{\mathbf{x}},$$

where we used the symmetry of A. Taking the complex-conjugate of the expression for z_1 we find

$$\overline{z_1} = \overline{(\mathbf{x}^T A\overline{\mathbf{x}})} = \overline{\mathbf{x}}^T A\mathbf{x} = z_1,$$

so that $z_1 \in \mathbb{R}$. Moreover,

$$z_2 = \overline{\mathbf{x}}^T\mathbf{x} = \sum_{i=1}^n \overline{x_i}x_i = \sum_{i=1}^n |x_i|^2 \in \mathbb{R}.$$

Hence $\lambda = z_1/z_2$ is real as the ratio of two real scalars.

ii. Let $\lambda \in \mathbb{R}$, and let $\mathbf{x} \in \mathbb{C}^n$ be the corresponding eigenvector. Let $\mathbf{x} = \mathbf{y} + i\mathbf{z}$ for some vectors $\mathbf{y}, \mathbf{z} \in \mathbb{R}^n$. Then both $\mathbf{y}$ and $\mathbf{z}$ must be eigenvectors associated with λ, since

$$A\mathbf{x} = \lambda\mathbf{x} \iff A(\mathbf{y} + i\mathbf{z}) = \lambda(\mathbf{y} + i\mathbf{z}) \iff \begin{cases} A\mathbf{y} = & \lambda\mathbf{y}, \\ A\mathbf{z} = & \lambda\mathbf{z}, \end{cases}$$

by comparing real and imaginary parts. Since eigenvectors are not unique, we can choose a convenient form, i.e., a real eigenvector for the real eigenvalue λ.

■

R Note that if $\mathbf{y}$ and $\mathbf{z}$ are real eigenvectors, then we must have $\mathbf{z} = c\mathbf{y}$ for some (real) scalar c. Then the complex eigenvector $\mathbf{x}$ satisfies $\mathbf{x} = (1 + ic)\mathbf{z}$, which is what we would expect: eigenvectors are unique up to multiplication by a scalar.

This result does not necessarily imply that the matrix is diagonalisable over $\mathbb{R}$; in general, this limited spectral information is insufficient to guarantee that alegbraic and geometric multiplicities are equal for each eigenvalue, i.e., that the matrix is not defective. We consider this next.

The following theorem is one of the major results of Linear Algebra. The statement can be equivalently given for symmetric matrices and self-adjoint maps. We include the latter version below.

Theorem 20.7 — Real Spectral Theorem. Let V be a real inner product space of dimension n and let $f \in \mathcal{L}(V)$ be self-adjoint. Then f is diagonalisable, with real eigenvalues and real eigenvectors which form an orthonormal basis for V.

Proof. By Lemma 20.6, the eigenvalues are real, so we only have to show that the matrix is diagonalisable, with orthonormal eigenvectors. We show this by induction on the dimension of V.

$P(n = 1)$: Let $n = 1$. Then the statement holds trivially.
$P(n = k-1)$: Assume that the statement holds for $n = k-1$: all self-adjoint maps on $(k-1)$-dimensional real inner product spaces are diagonalisable with real eigenvalues and orthonormal corresponding eigenvectors.
$P(n = k-1) \Rightarrow P(n = k)$: Let $n = k$ and consider a real eigenpair $(\mu, \mathbf{u})$ of f, where we assume that $\mathbf{u}$ is a unit vector. Let $U = \text{span}\{\mathbf{u}\}$ and consider its orthogonal complement in V: $U^\perp = \{\mathbf{v} \in V : \langle \mathbf{v}, \mathbf{u}\rangle = 0\}$. By Proposition 8.9, $V = U \oplus U^\perp$, so that dim $U^\perp = k-1$. Now, $U^\perp$ is an f-invariant subspace of V since for all $\mathbf{v} \in U^\perp$, there holds $f(\mathbf{v}) \in U^\perp$ as

$$\langle f(\mathbf{v}), \mathbf{u}\rangle = \langle \mathbf{v}, f(\mathbf{u})\rangle = \langle \mathbf{v}, \lambda\mathbf{u}\rangle = \mu \langle \mathbf{v}, \mathbf{u}\rangle = 0.$$

By Proposition 18.1, the restriction $\tilde{f}$ of f to $U^\perp$ is an endomorphism on $U^\perp$, i.e., $\tilde{f} \in \mathcal{L}(U^\perp)$. Moreover, $\tilde{f}$ is self-adjoint on $U^\perp$, since for all $\mathbf{v}_1, \mathbf{v}_2 \in U^\perp$

$$\left\langle \tilde{f}(\mathbf{v}_1), \mathbf{v}_2 \right\rangle = \langle f(\mathbf{v}_1), \mathbf{v}_2\rangle = \langle \mathbf{v}_1, f(\mathbf{v}_2)\rangle = \left\langle \mathbf{v}_1, \tilde{f}(\mathbf{v}_2)\right\rangle.$$

We conclude that $\tilde{f}$ is a self-adjoint endomorphism on a space of dimension $k-1$; by the inductive hypotesis it is diagonalisable, with $k-1$ real eigenvalues and $k-1$ real eigenvectors which form an orthonormal basis of $U^\perp$, say B_{k-1}. However, any eigenpair of $\tilde{f}$ is also an eigenpair of f, since $\lambda\mathbf{v} = \tilde{f}(\mathbf{v}) = f(\mathbf{v})$. Therefore the spectrum of f is $\text{sp} f = \{\mu\} \cup \text{sp}\tilde{f}$ and we note that these are all the eigenvalues of f. The corresponding eigenvectors form a set with n elements $B_k = \{\mathbf{u}\} \cup B_{k-1}$, which is linearly independent, as $\mathbf{u} \perp B_{k-1}$ where B_{k-1} is an orthonormal basis of $U^\perp$, by the inductive hypothesis. Hence B_k is an orthonormal basis of V. ■

We immediately obtain the following result, by using the results of the Spectral Theorem and of Theorem 20.5.

Corollary 20.8 Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then A is diagonalisable, with real eigenvalues and real eigenvectors which form an orthonormal basis for $\mathbb{R}^n$.

We can immediately derive the canonical form for the matrix representation of a self-adjoint endomorphism:

$$A = QDQ^T,$$

where

- D is a diagonal matrix containind the real eigenvalues of A;
- Q is a matrix with orthonormal vectors; this is known as an orthogonal matrix; we note here that orthonormality implies $Q^TQ = QQ^T = I_n$ so that $Q^{-1} = Q^T$.

This representation of A is also known as the spectral decomposition of A.

R A similar description can be derived for the complex case. For details, see the references provided.