

LECTURE 14

Endomorphisms

In this lecture we consider linear maps from a vector space to itself: $f: V \rightarrow V$, 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 $\mathbb{F} = \mathbb{C}$.

14.1 Endomorphisms

Definition 14.1 A linear map $f: V \rightarrow V$ 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 $\text{End}(V)$.

What is special about the domain and codomain being equal? Here are some observations.

Proposition 14.1 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 13.1. ■

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 $f: V \rightarrow V$ 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 14.2 — 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 14.2 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 14.1 Let us revisit a previous example: the differentiation map. However, this time we define it as an endomorphism: $D: V \rightarrow V$, with $V = \mathcal{P}_n(\mathbb{R})$ and $(Dp)(x) = p'(x)$. Let $B_V = \{q_j = x^{j-1}, j = 1, 2, \dots, n+1\}$. Then the matrix representation $A \in \mathbb{R}^{(n+1) \times (n+1)}$ satisfies for all $j = 1, \dots, n+1$

$$D(q_j) = \sum_{i=1}^{n+1} a_{ij} q_i.$$

Let us consider the case $j = 1$ first:

$$j = 1 : 0 = D(q_1) = \sum_{i=1}^{n+1} a_{i1} q_i \implies a_{i1} = 0, \quad i = 1, 2, \dots, n+1.$$

For any $j > 1$, matching the monomials on the left and right, we find

$$D(q^j) = (j-1)x^{j-2} = \sum_{i=1}^{n+1} a_{ij} x^{i-1} \iff a_{ij} = \begin{cases} 0 & j-2 \neq i-1 \\ j-1 & j-2 = i-1 \end{cases} \iff a_{ij} = \begin{cases} 0 & i \neq j-1, \\ j-1 & i = j-1. \end{cases}$$

The resulting matrix is included below.

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & n \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{(n+1) \times (n+1)}.$$

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 14.2 — Projection map. The projection map is a map $\pi: V \rightarrow V$ 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.

$$\begin{array}{ccccccc} \mathbf{v} & \xrightarrow{\pi} & \pi(\mathbf{v}) & \xrightarrow{\pi} & \pi^2(\mathbf{v}) = \pi(\mathbf{v}) \\ \varphi_V \downarrow & & \downarrow \varphi_V & & \downarrow \varphi_V \\ \mathbf{x} & \xrightarrow{P} & P\mathbf{x} & \xrightarrow{P} & P^2\mathbf{x} = P\mathbf{x} \end{array}$$

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.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \xrightarrow{\pi} \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}, \quad P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Note that in this case also the matrix representation is singular.

14.2 Automorphisms

Definition 14.3 An invertible endomorphism $f: V \rightarrow V$ is called an **automorphism**.

Automorphisms can also be seen as isomorphisms from a vector space to itself. We denote the set of automorphisms as $\text{Aut}(V)$. The following statements are immediate. We leave the proofs as an exercise.

Proposition 14.3 A linear transformation $f: V \rightarrow V$ is an automorphism if and only if $\ker f = \{\mathbf{0}_V\}$.

Corollary 14.4 A linear transformation $f: V \rightarrow V$ is an automorphism if and only if f has full rank.

Proposition 14.5 The inverse of an automorphism is an automorphism.

Proposition 14.6 $\text{Aut}(V)$ is not a subspace of $\mathcal{L}(V)$.

(R) While $\text{Aut}(V)$ fails to represent an algebraic structure with respect to standard function addition, it

does form an algebraic structure with respect to function composition. Since existence of an inverse is satisfied with respect to composition, $(\text{Aut}(V), \circ)$ is a non-Abelian group.

The matrix representations of maps $f \in \text{Aut}(V)$ are invertible, since all f are invertible (see Proposition 13.3). 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 14.7 Let $f : V \rightarrow U, g : U \rightarrow W$. Then

$$\text{rank } g \circ f \leq \min \{\text{rank } f, \text{rank } g\}.$$

Proof. We have

$$\text{im } g \circ f = g(f(V)) \subseteq g(U) = \text{im } g \implies \text{rank } g \circ f \leq \text{rank } g.$$

To obtain the other bound, we define the restriction map $\tilde{g} : \text{im } f \rightarrow W$, 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 14.8 Let $f \in \mathcal{L}(V, W)$. Let $g_1 \in \text{Aut}(V)$ and $g_2 \in \text{Aut}(W)$. Then

$$\text{rank } g_1 \circ f = \text{rank } f, \quad \text{rank } f \circ g_2 = \text{rank } f.$$

Proof. Let $r_1 = \text{rank } g_1 \circ f$. Then, but Proposition 14.7,

$$r_1 \leq \min \{\text{rank } g_1, \text{rank } f\} \implies r_1 \leq \text{rank } f.$$

Since $g_1 \in \text{Aut}(V)$, g_1 is invertible. Then, using again Proposition 14.7,

$$\text{rank } f = \text{rank } g_1^{-1} \circ (g_1 \circ f) \leq \min \{\text{rank } g_1^{-1}, \text{rank } g_1 \circ f\} \implies \text{rank } f \leq \text{rank } g_1 \circ f \implies \text{rank } f \leq r_1.$$

Hence, we have $r_1 = \text{rank } f$. The second statement follows similarly. ■

14.3 Canonical forms

Let us recall the canonical form obtained for the case of linear maps $f : V \rightarrow W$ (see Proposition 12.5):

$$A^{\text{can}} = \begin{bmatrix} I_r & O_{r,n-r} \\ O_{m-r,r} & O_{m-r,n-r} \end{bmatrix}.$$

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 $f : V \rightarrow V$, 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 14.4 — 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

$$AM = MB \implies A\mathbf{c}_i(M) = b_{ii}\mathbf{c}_i(M) \quad (i = 1, 2, \dots, n),$$

where $\mathbf{c}_i(M)$ denotes the i th column of M . This is in fact a familiar relation, particularly when written in the form

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i \quad (i = 1, 2, \dots, n).$$

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 14.5 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 $B = M^{-1}AM$, with 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 14.3 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 way¹. 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.

14.3.1 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

$$U^T AV = \Sigma,$$

¹We will see later that symmetric matrices are diagonalisable.

where $\Sigma \in \mathbb{R}^{m \times n}$ is the diagonal matrix

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_n & \\ \hline & & & \mathbf{0}_{(m-n) \times n} \end{bmatrix},$$

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 .



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

$$Q^T A Q = T = D + U,$$

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

$$Q^T A Q = H,$$

where H is a Hessenberg matrix, i.e., an upper triangular matrix with an additional sub-diagonal.