LECTURE 11

Matrix representations (2)

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.

11.1 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 11.1 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 $x_i = [\mathbf{x}]_i, i = 1, 2, \ldots, n$ 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:

$$B_V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}, \qquad B_W = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}.$$

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 11.2 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 11.3 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 9.6 and also using the above corollary,

$$\dim\ \ker f = |B_0| = \dim\ \varphi_V(\mathrm{span} B_0) = \dim\ \varphi_V(\ker f) = \dim\ \ker A. \quad \blacksquare$$

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 11.1 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 $A\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \iff \begin{cases} x_3 & = 0 \\ x_1 + 2x_2 & = 0 \end{cases} \iff \ker A = \mathrm{span}\left\{ \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} \right\}.$$

Hence, $\dim\ \ker f = 1$, with $\ker f = \mathrm{span}\{\mathbf{v}_0\}$, 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:

$$f(p) = 0 \iff a_2 + (a_0 + a_1 + a_2)\mathbf{x} = 0 \iff \begin{cases} a_2 & = 0 \\ a_0 + a_1 & = 0 \end{cases} \iff \ker f = \mathrm{span}\{p_0 := 1 - x\}.$$

However, in practice, it is more straightforward to implement (on a computer) the computation of $\ker A$.

Proposition 11.4 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

$$\mathbf{w} \in \text{im} f \iff \mathbf{w} = f(\mathbf{v}) \iff \varphi_W(\mathbf{w}) = \varphi_W(f(\mathbf{v})) \iff \mathbf{y} = (\varphi_W \circ f)(\mathbf{v}) = (\alpha \circ \varphi_V)(\mathbf{v}) = A\mathbf{x} \iff \mathbf{y} \in \ker A.$$

■

We immediately derive the following corollary.

Corollary 11.5 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 11.6 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 9.6 and also using the above corollary,

$$\text{rank } f = \dim \text{ im } f = |B| = \dim\, \varphi_W(\text{span} B) = \dim\, \varphi_W(\text{im } f) = \dim \text{ col } A = \text{rank } A.$$

■

An immediate consequence of the above results and the rank-nullity formula is the following.

Corollary 11.7 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 11.1 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.

11.2 Matrix representation of compositions of linear maps

Let us recall, as well as introduce, some definitions, results and notation for composition of functions.

Definition 11.1 — 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 11.8 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 11.9 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 $C = AB$.

(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 10.1, the action of h on

[1] See Definition 5.1.

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 11.10 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.