Lecture IV : Matrices & Linear Transformations

If $T : V \to W$ is a linear transformation from a finite dimensional vector space V to a finite dimensional vector space W and if $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ are bases for V and W respectively then T is determined by knowing the double indexed set α_i^j of numbers determined by

$$T(v_i) = \sum_{j=1}^{m} \alpha_i^j w_j \qquad \begin{array}{c} i = 1, \ldots n \\ (j = 1, \ldots m). \end{array}$$

It is conventional to regard this double indexed set as arranged in a rectangle a "matrix". We adopt the specific convention that each column of the rectangle is the ordered m-tuple of components of the W-vector $T(v_i)$, $i = 1, \ldots n$. So i labels which column, j labels which row. And following usual notation we write the matrix as $a_{ij} = \alpha_j^i$

so i is the row index, j is the column index (note "interchange" in a sense of i and j)

Namely

$$T(v_i) = \sum_{j=1}^{m} a_{ji} w_j \qquad i = 1, \ldots n.$$

The initially slightly odd-looking choice of row versus column indexing has the good property that

$T\left(\sum_{i=1}^{n} \beta_j v_i\right)$ written as a column vector

of $w_1, \ldots, w_m$ components (m of them)

$= \beta_1 \begin{pmatrix}\text{column } 1 \\ \text{of matrix}\end{pmatrix} + \cdots + \beta_n \begin{pmatrix}\text{column } n \\ \text{of matrix}\end{pmatrix}$

$= A \begin{pmatrix}\beta_1 \\ \vdots \\ \beta_n\end{pmatrix}$ where the product is in the usual matrix product sense.

We recall what that sense (of usual product is): Given A_{ij}, $i = 1, \ldots, n_1$, $j = 1, \ldots, n_2$ (n_1 rows, n_2 columns) and B_{jk}, $j = 1, \ldots, n_2$, $k = 1, \ldots, n_3$ (n_2 rows, n_3 columns) then

$A \times B$ is a matrix with n_1 rows and n_3 columns defined by

$$(A \times B)_{ik} = \sum_{j=1}^{n_2} A_{ij} B_{jk} \qquad \begin{matrix} i \in \{1, \ldots, n_1\} \\ k \in \{1, \ldots, n_3\} \end{matrix}$$

This is the usual "across A, down B" multiplication rule.

Computational exercise: Matrix multiplication is associative (formulation of sizes is part of the exercise!)

In the case of $T: V \to W$, V dimension n_1, W dimension n_2, the matrix of T has n_2 rows and n_1 columns, the domain column vector has n_1 rows and 1 column (n_1 rows are V-components), and the product has n_2 rows, 1 column (column of W components).

Consequence of associativity of matrix multiplication: If $T: V \to W$ and $S: W \to U$, V, W, U with bases (finite dimensional assume for all three!) then the matrix for T first, then S, linear transformation from V to U is

$B \times A$ if the matrix for T is A and the matrix for S is B.

Watch out for order! It matters.
E.g. $T: \mathbb{R}^2 \to \mathbb{R}^2$ $T((x,y)) = (0, y)$
$S((x,y)) = (y,x)$, $S: \mathbb{R}^2 \to \mathbb{R}^2$.
S first, then T takes (x,y) to $(0,x)$
T first, then S takes (x,y) to $(y,0)$.
These are different unless $(x,y) = (0,0)$!

Exercise: Write this example in matrix terms and compute the products.

Definition: If a matrix A has n_1 rows and n_2 columns, then the rows can be considered each a vector in $\mathbb{R}^{n_2}$. The columns can be considered as vectors in $\mathbb{R}^{n_1}$. The row rank of A is the dimension of the subspace of $\mathbb{R}^{n_2}$ spanned by the "row vectors". The column rank

= the dimension of the subspace in $\mathbb{R}^{n_1}$ spanned by the "column vectors".

It is an interesting and important fact that row rank = column rank.

We shall show this using the following:

Lemma: If $T: V \to W$ (V, W finite dimensional) is a linear transformation with matrix A relative to bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ for V and W respectively then $T^*: W^* \to V^*$ has matrix A^t relative to the bases $w_1^* \ldots w_m^*$ and $v_1^* \ldots v_n^*$ of W^* and V^* respectively where as usual the transpose of a matrix A_{ij}, written A^t is defined by

$$A^t_{\cdot\,\cdot\cdot} = A_{\cdot\cdot\,\cdot}$$

(Here the row index of A^t has the range of the column index of A and the column index of A^t has the range of the row index of A).

Example

$$\begin{pmatrix} 2 & 3 & 4 \\ 5 & 6 & 7 \end{pmatrix}^t = \begin{pmatrix} 2 & 5 \\ 3 & 6 \\ 4 & 7 \end{pmatrix}$$

"rows and columns are interchanged")

Proof: Let $\dim V = n$ and $\dim W = m$ as indicated, bases as indicated. Then the no of columns of the matrix B of $T^* : W^* \to V^*$ = dimension of $W^* = m$ and the number of rows = dimension of $V^* = n$. And the column vectors of B are the components of T^*(basis vector of W^*) in V^* basis. That is for the ith column of B (= image of w_i^*, $i \in \{1, .. m\}$), in the jth row there is the v_j^*-component of $T^*(w_i^*)$ $j \in \{1, .. n\}$, namely

$$B_{ji} = (T^*(w_i^*))(v_j)$$

The reason is that applying $L \in V^*$ to v_j^* "picks out" the jth component of $L = \sum_{l=1}^{n}$ (component for v_l^*) v_l^*.

So $\quad B_{ji} = (T^*(w_i^*))(v_j^*) \overset{\text{def on } T^*}{=} w_i^*(T(v_j))$

$=$ ith component of $T(v_j)$ expressed in W-basis $=$ A_{ij}, by definition. $\square$

Note that this is entirely formal and "definitional"!

Now if A is the matrix of $T: V \to W$ relative to bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$, then the column rank ~~of~~ of A =

dimension of $\operatorname{Im} T$. Since transpose interchanges rows and columns,

row rank of A = column rank of A^t
= dimension of $\operatorname{Im} T^*$

by the same logic. Thus (since we proved last lecture than $\dim(\operatorname{Im} T^*) = \dim(\operatorname{Im} T)$), every such A has row rank = column rank of A. But of course every matrix arises as the matrix of a linear transformation: just use the formula $T(v_i) = \sum a_{ji} w_j$ to define the linear transformation T. So

Theorem: For any matrix A,

row rank of A = column rank of A.

Invertible Linear Transformations & Matrices

Theorem:
If V is a finite dimensional vector space and $T: V \to V$ is a linear transformation then:
T injective $\iff$ T is surjective $\iff$ T is bijective.

Proof: $\dim \ker T + \dim \operatorname{im} T = \dim V$, so
$$\ker T = \{0\} \iff \operatorname{Im} T = V. \quad \square$$

Corollary: If $T: V \to V$ and $S: V \to V$ and T followed by S = identity map of V to V then T and S are both bijective.
Proof: S must be surjective. Hence S is injective. Also, T must be injective and hence surjective. $\square$

Translating this into the language of matrices (using a choice of basis for V)
Set I_n = matrix of identity map $V \to V$.
Theorem: If A and B are $n \times n$ matrices with $A \times B = I_n$, then A and B are the matrices of bijective linear transformations from V to V which are inverses of each other. In particular
$$B \times A = I_n.$$

Definition: An $n \times n$ matrix is <u>invertible</u> if there is a matrix, also $n \times n$, such that $A \times B = I_n$.

This happens if and only if there is a matrix, also $n \times n$, C such that $C \times A = I_n$. And $C = B$ (if either and hence both exist).

Def: The matrix B (or C) as above is the <u>inverse</u> of A, denoted A^{-1}.

How can we decide, given an $n \times n$ matrix A, whether it is invertible or not?
First, we consider A as determining a linear transformation T_A from F^n to F^n (F = the field of choice). Then T_A is surjective if and only if the columns of A are linearly independent, thought of as vectors in F^n, since the columns are the T_A images of the "standard basis" vectors of F^n.
With this in mind, we can characterize the invertibility or not of A as follows:

Theorem: A is invertible if and only if $\det A \neq 0$.

Proof: From homework assignment I, linear dependence of the columns of $A \Rightarrow \det A = 0$ (problem): This is straightforward.

Less obvious is problem , homework I:

If the columns of A are independent, then $\det A \neq 0$: this is done by "column operations" which preserve whether A has nonzero determinant or zero determinant. When A has independent columns, such operations turn A into a matrix of the form

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ * & 1 & 0 & \cdots \\ * & * & 1 & \ddots \\ & & & \ddots & 1 \end{pmatrix}$$

(all 0's above main diagonal) which clearly has determinant 1. Details are left as an exercise.

Since we already know that A is invertible $\Leftrightarrow$ columns of A are independent, the Theorem follows. $\square$.

Corollary: The set of invertible $n \times n$ matrices, considered as a subset of $\mathbb{R}^{n^2}$, is open.

Proof: $\det A$ is clearly a continuous function of the entries of A: it is a polynomial. So $\{A : \det A \neq 0\}$ is open. $\square$

It is of interest to establish the corollary from another viewpoint because this alternative viewpoint can be applied in the infinite dimensional cases on some occasions (which determinants cannot, be by nature finite-dimensional items, at least as we have defined them).

The idea is as follows:
Given T_0 an invertible linear transformation, we want to know that $T_0 + T_1$ is invertible if T_1 is small enough. Since $T_0 + T_1$

$= T_0(I + T_0^{-1}T_1)$, it is enough to show that $I_n + A$ is invertible when A is small (T_1 small makes $T_0^{-1}T_1$ small, if T_0 is fixed). Now, in analogy with

$1/(1+r) = 1 - r + r^2 - r^3 \ldots, \quad |r| < 1$, we might hope that (when A is small)

$(I+A)^{-1} = I - A + A^2 - A^3 \ldots$
On the formal level, this works: powers of A commute and

$I \times I - A + A^2 \ldots = I - A + A^2 - A^3 \ldots$
$A \times (I - A + A^2 \ldots) = A - A^2 + A^3 \ldots$

So
$(I+A)(I - A + A^2 - A^3 \ldots) = I$.
The question is how to get some convergence ideas to make this work in a precise sense.

Definition: If $A: \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation, then the <u>operator</u> <u>norm</u> $\|A\|$ is by definition

$$\sup\{\|Av\| : v \in \mathbb{R}^n, \|v\| = 1\}$$

when $\| \; \|$ of $\mathbb{R}^n$ elements are the usual

$$\|(x_1, \ldots, x_n)\| = \left(\sum |x_i|^2\right)^{1/2}.$$

By standard analysis (continuous function on compact set), $\|A\|$ is finite.
It is easy to check that $\|A \times B\|$

$\leq \|A\| \|B\|$ for any two matrices (or $A \times B$ considered as composition of linear transformations)
So
$\|A^n\| \leq \|A\|^n$. Also $\|A_1 + A_2\| \leq \|A_1\| + \|A_2\|$
In particular, if $\|A\| < 1$, then $I - A + A^2 - A^3 \ldots$ converges in operator norm (more detail: it is Cauchy in operator norm, and matrices / linear

transformations are a complete metric space in $\operatorname{dis}(A,B) = \|A-B\|$, this metric being equivalent to the $\mathbb{R}^{n^2}$ one).

Convergence being established,

$$(I+A)(I-A+A^2-A^3\ldots) = I$$

follows easily since

$$(I+A)(I-A+A^2-A^3\ldots \pm A^N)$$

$$= \pm A^{N+1}$$

and $\|A^{N+1}\| \to 0$ as $N \to +\infty$. (when $\|A\|<1$). $\square$

In the infinite-dimensional cases, one has to assume that the sup used to define the operator norm (when the transformations are operating on a vector space which itself has a norm) are bounded: details of this are left for later (if ever!)