Math 350: Linear Algebra

Wolmer V. Vasconcelos

Set 3

Spring 2010

Big Picture

$$\begin{aligned}\textbf{Linear Algebra} &= \textbf{Vector Spaces}\\ &+ \textbf{Linear Transformations ++}\\ &+ \textbf{Applications}\end{aligned}$$

Outline

Functions on Vector Spaces

Let **V** and **W** be two vector spaces over the field **F**. What are the functions like between these spaces?:

$$\mathbf{T} : \mathbf{V} \to \mathbf{W}.$$

V is called the source, and **W** the target of the function. For example, suppose $\mathbf{V} = \mathbf{W} = \mathbf{F}^2$. Then **T** takes for input pairs $v = (x_1, x_2)$, and outputs pairs $\mathbf{T}(v) = (y_1, y_2)$:

$$(x_1, x_2) \to \boxed{\quad \mathbf{T} \quad} \to (y_1, y_2) = (\mathbf{f}_1(x_1, x_2), \mathbf{f}_2(x_1, x_2))$$

It can be very varied since functions of two variables come in many flavors.

We will be looking at certain type of functions illustrated by the following examples.

- Let **V** be the vector space of all real valued functions with derivatives in $[-1, 1]$, and let **W** be the vector space of real valued functions on $[-1, 1]$. Define

$$\mathbf{T}(f(t)) = f'(t),$$

  or

$$\mathbf{L}(f) = f'' - f.$$

- Here are two other functions

$$\mathbf{T}(f) = \int_{-1}^{1} f(t)dt, \quad \mathbf{T} : \mathbf{V} \to \mathbb{R}$$

$$\mathbf{L}(f) = \int_{-1}^{t} f(t)dt, \quad \mathbf{T} : \mathbf{V} \to \mathbf{W}$$

- $\mathbf{T} : \mathbb{R}^2 \to \mathbb{R}^2$

$$\mathbf{T}(x, y) = (y, x)$$

  This is reflection about the [main] diagonal.

- For α fixed,

$$\mathbf{T}(x, y) = (x \cos\alpha + y \sin\alpha, -x \sin\alpha + y \cos\alpha)$$

  This is a rotation in the plane by α degrees.

- $\mathbf{T} : \mathbb{R}^3 \to \mathbb{R}^2$

$$\mathbf{T}(x, y, z) = (x, y)$$

  This is projection on the *xy*-plane.

All these functions share the following property:

Definition

A function $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ is a linear transformation, or linear operator, if it satisfies:

(i) For any $v_1, v_2 \in \mathbf{V}$,

$$\mathbf{T}(v_1 + v_2) = \mathbf{T}(v_1) + \mathbf{T}(v_2)$$

[**T** is additive, that is takes sums to sums]

(ii) For any $v \in \mathbf{V}$ and $c \in \mathbf{F}$,

$$\mathbf{T}(cv) = c\mathbf{T}(v)$$

[**T** commutes with scaling]

We can put these two properties together:
A function can be viewed as a factory processing inputs into outputs

$$\text{input} \longrightarrow \boxed{\mathbf{f}} \longrightarrow \text{output}$$

One key property of a linear box is that it can be reverse engineered.

Proposition

Let $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ *be a linear transformation of vector spaces. If* $v_1, \ldots, v_n \in \mathbf{V}$ *and* $c_1, \ldots, c_n \in \mathbf{F}$*, then*

$$\mathbf{T}(\sum_{i=1}^{n} c_i v_i) = \sum_{i=1}^{n} c_i \mathbf{T}(v_i).$$

[That is, $\mathbf{T}$ *commutes with linear combinations.]*

Proof.

It uses the conditions (i) and (ii) of the definition and induction: Apply $\mathbf{T}$ to

$$\sum_{i=1}^{n} c_i v_i = (\sum_{i=1}^{n-1} c_i v_i) + c_n v_n$$

and iterate. □

Recipe for linear transformations

Let **V** be a vector space with a basis $v_1, v_2, \ldots$. If **W** is a vector space, for each v_i choose $w_i \in \mathbf{W}$ [the w_i do not need to be linearly independent].

Proposition

The assignment

$$\sum_i x_i v_i \mapsto \sum_i x_i w_i$$

defines a linear transformation from **V** *to* **W**.

One quick way to build a L.T. between spaces of tuples is the following. Let $\mathbf{A}$ be an $m \times n$ matrix with entries in the field $\mathbf{F}$. For a n-tuple

$$v = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

define the function $\mathbf{L_A} : \mathbf{F}^n \to \mathbf{F}^m$

$$\mathbf{L_A}(v) = \mathbf{A} \cdot v.$$

Since multiplication of matrices is distributive and commutes with scaling, $\mathbf{L_A}$ is a L.T.

Let $\mathbf{M}_2(\mathbf{F})$ be the vector space of all 2-by-2 matrices over the field $\mathbf{F}$.
Fix a matrix, say, $\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}$ and define the function

$$\mathbf{B} \to \mathbf{T}(\mathbf{B}) = \mathbf{AB}.$$

$\mathbf{T}$ satisfies:

$$\begin{aligned} \mathbf{T}(\mathbf{B}_1 + \mathbf{B}_2) &= \mathbf{A}(\mathbf{B}_1 + \mathbf{B}_2) = \mathbf{AB}_1 + \mathbf{AB}_2 \\ &= \mathbf{T}(\mathbf{B}_1) + \mathbf{T}(\mathbf{B}_2) \\ \mathbf{T}(c\mathbf{B}) &= \mathbf{A}(c\mathbf{B}) = c\mathbf{AB} = c\mathbf{T}(B). \end{aligned}$$

Point: Lots of freedom to create linear transformations.

There are several subsets associated to a linear transformation $\mathbf{T} : \mathbf{V} \to \mathbf{W}$:

- The Nullspace or Kernel of $\mathbf{T}$ is the subset

$$N(\mathbf{T}) = \{v \in \mathbf{V} \mid \mathbf{T}(v) = O.\}$$

  [The vectors mapped to O]
- The Range or Image of $\mathbf{T}$ is the subset

$$R(\mathbf{T}) = \{w \in \mathbf{W} \mid w = \mathbf{T}(v), \quad v \in \mathbf{V}.\}$$

Examples

If **T** is the linear transformation

$$f \mapsto f'' - f$$

defined earlier, its nullspace consists of the solutions of $y'' - y = 0$, that is the linear combinations

$$ae^x + be^{-x}, \quad a, b \in \mathbb{R}.$$

Proposition

The <u>*Nullspace*</u> *and the* *Range* *of a linear transformation* $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ *are subspaces of* $\mathbf{V}$ *and* $\mathbf{W}$ *respectively.*

Proof.

Let us apply the subspace test to $N(\mathbf{T})$. Suppose $v_1, v_2 \in N(\mathbf{T})$. Then for any scalars c_1, c_2,

$$\mathbf{T}(c_1 v_1 + c_2 v_2) = c_1 \mathbf{T}(v_1) + c_2 \mathbf{T}(v_2) = c_1 O + c_2 O = O.$$

So the linear combination belongs to the Nullspace.
We leave for you the other proof. □

The dimension of $N(\mathbf{T})$ is called the nullity and the dimension of $R(\mathbf{T})$ is called the rank of $\mathbf{T}$.

Dimension Formula

Theorem

Let $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ *be a linear transformation of finite dimensional vector spaces. Then*

$$\dim N(\mathbf{T}) + \dim R(\mathbf{T}) = \dim \mathbf{V}.$$

That is, *nullity* $+$ *rank* $= \dim \mathbf{V}$*.*

Proof. Suppose $v_1, \ldots, v_n$ is a basis of $\mathbf{V}$, and $u_1, \ldots, u_r$ is a basis of $N(\mathbf{T})$. We are going to show that $R(\mathbf{T})$ has basis with $n - r$ elements. Recall that $\mathbf{T}(\sum_{i=1}^n c_i v_i) = \sum_{i=1}^n c_i \mathbf{T}(v_i)$, $R(\mathbf{T})$ is spanned by

$$\mathbf{T}(v_1), \mathbf{T}(v_2), \ldots, \mathbf{T}(v_j), \ldots, \mathbf{T}(v_n).$$

Out of this list we are going to pick a basis for $R(\mathbf{T})$.

We scan the list and delete the vectors [red] that can be written as linear combination of the preceding vectors

$$\mathbf{T}(v_1), \mathbf{T}(v_2), \ldots, \mathbf{T}(v_j), \ldots, \mathbf{T}(v_n).$$

For convenience of notation we assume we are left with the first s vectors

$$\mathbf{T}(v_1), \mathbf{T}(v_2), \ldots, \mathbf{T}(v_s).$$

Claim: $u_1, \ldots, u_r, v_1, \ldots, v_s$ is a basis of **V**.
Once we have shown this we be done since all bases of **V** have n elements. Let us check.

Claim 1: $u_1, \ldots, u_r, v_1, \ldots, v_s$ spans $\mathbf{V}$

If $v \in \mathbf{V}$, $\mathbf{T}(v) = \sum_{i=1}^{s} a_i \mathbf{T}(v_i)$, that is

$$\mathbf{T}(v - \sum_{i=1}^{s} a_i v_i) = 0$$

that is $v - \sum_{i=1}^{s} a_i v_i$ belongs to the nullspace so

$$v - \sum_{i=1}^{s} a_i v_i = \sum_{j=1}^{r} b_j u_j.$$

Claim 2: $u_1, \ldots, u_r, v_1, \ldots, v_s$ are linearly independent.

If $\sum b_j u_j + \sum a_i v_i = O$, applying $\mathbf{T}$ we get $\sum a_i \mathbf{T}(v_i) = 0$ since $\mathbf{T}(u_j) = 0$. But the $\mathbf{T}(v_i)$ are linearly independent [they form a basis of $R(\mathbf{T})$] so $a_i = 0$. We have left $\sum b_j u_j = 0$, which implies $b_j = 0$ since the u_j form a basis of $N(\mathbf{V})$

Let us recall some general properties of a function $\mathbf{f} : \mathbf{X} \to \mathbf{Y}$

- $\mathbf{f}$ is one-one if $\mathbf{f}(x_1) = \mathbf{f}(x_2)$ implies $x_1 = x_2$. One also says that $\mathbf{f}$ is injective. If $\mathbf{f}$ is a linear transformation, $\mathbf{f}(x_1) = \mathbf{f}(x_2)$ means $\mathbf{f}(x_1 - x_2) = O$ so $\mathbf{f}$ is one-one if and only if the nullspace is (O).

- $\mathbf{f}$ is onto if its image is $\mathbf{Y}$. One also says that $\mathbf{f}$ is surjective.

- $\mathbf{f}$ is an isomorphism or invertible when it is both.

Here are some consequences of the dimension formula applied to a linear transformation $\mathbf{T} : \mathbf{V} \to \mathbf{W}$

- If $\dim \mathbf{V} > \dim \mathbf{W}$, then $\mathbf{T}$ is not one-one
- If $\dim \mathbf{V} < \dim \mathbf{W}$, then $\mathbf{T}$ is not onto.
- If $\dim \mathbf{V} = \dim \mathbf{W}$, then $\mathbf{T}$ is an isomorphism iff its nullspace is O, or iff is onto.

Exercise: If $\mathbf{T} : \mathbf{V} \rightarrow \mathbf{V}$ and $\mathbf{T}^2 = 0$, show that nullity $\geq$ rank.

Answer: Clear since all vectors of the range are like $\mathbf{T}(v)$, but these vectors $\mathbf{T}(\mathbf{T}(v)) = \mathbf{T}^2(v) = O$.

Thus the **range** of **T** is contained in its **nullspace**, and therefore

$$\text{nullity} \ = \dim(\text{nullspace}) \geq \dim(\text{range}) \ = \text{rank}$$

Exercise 1a: Section 2.1, 19, 25.

Let us use these ideas to solve anew an earlier exercise: If $\mathbf{S}_1$ and $\mathbf{S}_2$ are subspaces of a V.S. $\mathbf{V}$, then $\dim \mathbf{S}_1 + \dim \mathbf{S}_2 = \dim(\mathbf{S}_1 + \mathbf{S}_2) + \dim(\mathbf{S}_1 \cap \mathbf{S}_2)$. Consider the mapping $\mathbf{T} : \mathbf{S}_1 \times \mathbf{S}_2 \to \mathbf{V}$, given by

$$\mathbf{T}(u, v) = u + v.$$

This clearly a L.T. The range $R(\mathbf{T})$ of $\mathbf{T}$ is the subspace $\mathbf{S}_1 + \mathbf{S}_2$. What is its nullspace $N(\mathbf{T})$?It consists of the pairs (u, v) with $u + v = O$. That is, the elements of the form $(u, -u)$ with $u \in \mathbf{S}_1 \cap \mathbf{S}_2$. This implies that $N(\mathbf{T})$ is isomorphic to $\mathbf{S}_1 \cap \mathbf{S}_2$, in particular have the same dimension. Since $\dim(\mathbf{S}_1 \times \mathbf{S}_2) = \dim \mathbf{S}_1 + \dim \mathbf{S}_2$, we are done.

Outline

Matrix Representation

We first discuss how to represent some [look at the caveat] linear transformations $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ by matrices. Think of $\mathbf{V}$ and $\mathbf{W}$ as $\mathbb{R}^n$ or $\mathbb{C}^n$. It is a process akin to representing vectors by coordinates. Recall that if $v \in \mathbf{V}$ and $\mathcal{B} = v_1, \ldots, v_n$ is a basis of $\mathbf{V}$, we have a unique expression

$$v = x_1 v_1 + \cdots + x_n v_n.$$

We say that the x_i are the coordinates of v with respect to $\mathcal{B}$. We write as

$$[v]_{\mathcal{B}} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

If $\mathcal{C} = \{w_1, \ldots, w_m\}$ is a basis of $\mathbf{W}$, we would like to find the coordinates of $\mathbf{T}(v)$ in the basis $\mathcal{C}$

$$[\mathbf{T}(v)]_{\mathcal{C}} = \begin{bmatrix} ? \end{bmatrix}.$$

In other words, if $v = x_1 v_1 + \cdots + x_n v_n$,

$$\mathbf{T}(v) = y_1 w_1 + \cdots + y_m w_m,$$

we want to describe the y_i in terms of the x_j. The process will be called a matrix representation. It comes about as follows:

$$\sum y_i w_i = T(\sum x_j v_j) = \sum x_j \mathbf{T}(v_j)$$

Thus if we have the coordinates of the $\mathbf{T}(v_j)$,

$$\mathbf{T}(v_j) = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{bmatrix}$$

we have

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \sum x_j \begin{bmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{bmatrix}$$

More pictorially

$$[\mathbf{T}(v)]_{\mathcal{C}} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = [\mathbf{T}]_{\mathcal{B}}^{\mathcal{C}} \cdot [v]_{\mathcal{B}}$$

The $n \times m$ matrix

$$[\mathbf{T}]_{\mathcal{B}}^{\mathcal{C}}$$

is called the matrix representation of **T** relative to the bases $\mathcal{B}$ of **V** and $\mathcal{C}$ of **W**.

Quickly: Once bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ have been chosen, $\mathbf{T}$ is represented by

$$\left[\, a_{ij} \,\right]$$

where the entries come from

$$\mathbf{T}(v_j) = \sum_{i=1}^{m} a_{ij} w_i.$$

Example

Recall the transpose operation on a square matrix **A**: if a_{ij} is the (i, j)-entry of **A**, the (i, j)-entry of $\mathbf{A}^t$ is a_{ji}. This is a linear transformation **T** on the space $\mathbf{M}_n(\mathbf{F})$:

$$(\mathbf{A} + \mathbf{B})^t = \mathbf{A}^t + \mathbf{B}^t, \quad (c\mathbf{A})^t = c\mathbf{A}^t.$$

Let us find its matrix representation on $\mathbf{M}_2(\mathbf{F})$. This space has the basis

$$v_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, v_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, v_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, v_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Since

$$\mathbf{T}(v_1) = v_1, \quad \mathbf{T}(v_2) = v_3, \quad \mathbf{T}(v_3) = v_2, \quad \mathbf{T}(v_4) = v_4,$$

the matrix representation of transposing is

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

Example

Let $\mathbb{R}_3[x]$ be the space of real polynomials of degree at most 3 and $\mathbf{T}$ the differentiation operator.

A basis here are the polynomials $p_1 = 1, p_2 = x, p_3 = x^2, p_4 = x^3$. The corresponding matrix representation is

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{T}(p_4) = 3p_3$$

Note that the coordinates of $\mathbf{T}(p_4)$, $(0, 0, 3, 0)$ goes into the fourth column of the natrix representation.

Exercise

Suppose a linear transformation $\mathbf{T} : \mathbb{R}^3 \to \mathbb{R}^2$ satisfies

$$\mathbf{T}(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \mathbf{T}(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}) = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \quad \mathbf{T}(\begin{bmatrix} 1 \\ 4 \\ 9 \end{bmatrix}) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

(a) Show that the three vectors of $\mathbb{R}^3$ are linearly independent.
(b) Find the nullspace of this linear transformation.
(c) Find $\mathbf{T}(\begin{bmatrix} 1 \\ 0 \end{bmatrix})$.

Exercise 3: Section 2.2: 11.

Solution

Blocks

Suppose **T** is a L.T. of vector space **V** with a basis $\mathcal{A} = v_1, \dots, v_r, v_{r+1}, \dots, v_n$. Suppose $\mathbf{T}(v_i)$ for $i \le r$, is a linear combination of the first r basis vectors, and $\mathbf{T}(v_i)$ for $i > r$, is a linear combination of the last $n - r$ basis vectors.

To illustrate, suppose $\mathcal{A} = v_1, v_2, v_3, v_4$, and $\mathbf{T}(v_1) = av_1 + cv_2, \mathbf{T}(v_2) = bv_1 + dv_2$, and $\mathbf{T}(v_3) = ev_3 + fv_4, \mathbf{T}(v_3) = gv_3 + hv_4$, then the matrix representation of **T** for this basis is

$$[\mathbf{T}]_{\mathcal{A}} = \begin{bmatrix} a & c & 0 & 0 \\ b & d & 0 & 0 \\ 0 & 0 & e & f \\ 0 & 0 & g & h \end{bmatrix}$$

Claim: The matrix representation has the block format

$$[\mathbf{T}]_{\mathcal{A}} = \begin{bmatrix} \boxed{r \times r} & O \\ O & \boxed{(n-r) \times (n-r)} \end{bmatrix}$$

This can be refined to more than two blocks. The extreme case is when all blocks are 1×1. The representation is then said to be diagonal. If and when this happens is a major theme of Linear Algebra.

Addition of linear transformations

We are now going to combine linear transformations in various ways. Let $\mathbf{T}$ and $\mathbf{U}$ be two linear transformations of source $\mathbf{V}$ and target $\mathbf{W}$. Consider the operations,

$$\begin{aligned} (\mathbf{T}+\mathbf{U})(v) &:= \mathbf{T}(v)+\mathbf{U}(v) \\ (c\mathbf{T})(v) &:= c\mathbf{T}(v). \end{aligned}$$

Clearly they define [write the reasons] a vector space on the set $\mathcal{L}(\mathbf{V},\mathbf{W})$ of all such linear transformations.

Theorem

If $\mathbf{V}$ has dimension n and $\mathbf{W}$ has dimension m, then

$$\dim \mathcal{L}(\mathbf{V},\mathbf{W}) = m\cdot n.$$

We are going to build a basis for this space. Let $\mathcal{B} = v_1, \ldots, v_n$ be a basis of **V** and $\mathcal{C} = w_1, \ldots, v_m$ be a basis of **W**. Using the basic recipe, define the linear transformation

$$\mathbf{E}_{ij}(v_k) = \begin{cases} 0, & k \neq i \\ w_j, & k = i \end{cases}$$

There are $m \cdot n$ such [elementary] linear transformations.

Proposition

The $\mathbf{E}_{ij}$ *are linearly independent. [Which also follows from the next assertion.] If* $\mathbf{T}$ *is a linear transformation and*

$$\mathbf{T}(v_j) = \sum_i a_{ij} w_j,$$

then

$$\mathbf{T} = \sum_{i,j} a_{ij} \mathbf{E}_{ij}.$$

Proof. Try yourself or look up in book.

Exercise 4: Section 2.3: 11, 13.

Composition of linear transformations

There is another way to combine certain linear transformations. Consider composition of functions

$$\mathbf{V} \xrightarrow{\mathbf{T}} \mathbf{W} \xrightarrow{\mathbf{U}} \mathbf{Z},$$

$$(\mathbf{U} \circ \mathbf{T})(v) := \mathbf{U}(\mathbf{T}(v))$$

Proposition

With $\mathbf{T}$ *and* $\mathbf{U}$ *as above,* $\mathbf{U} \circ \mathbf{T}$ *is a linear transformation from* $\mathbf{V}$ *to* $\mathbf{Z}$.

Proof.

Let us check the basic requirements:

$$\begin{aligned}\mathbf{U}\circ\mathbf{T}(v_1+v_2) &:= \mathbf{U}(\mathbf{T}(v_1+v_2)) = \mathbf{U}(\mathbf{T}(v_1)+\mathbf{T}(v_2))\\ &= \mathbf{U}(\mathbf{T}(v_1))+\mathbf{U}(\mathbf{T}(v_2))\\ &= \mathbf{U}\circ\mathbf{T}(v_1)+\mathbf{U}\circ\mathbf{T}(v_2).\end{aligned}$$

It shows composition is additive.

$$\begin{aligned}\mathbf{U}\circ\mathbf{T}(cv) &:= \mathbf{U}(\mathbf{T}(cv)) = \mathbf{U}(c\mathbf{T}(v))\\ &= c\mathbf{U}(\mathbf{T}(v)) = c(\mathbf{U}\circ\mathbf{T})(v).\end{aligned}$$

It shows the scaling property. □

Now we are going to explain where multiplication of matrices comes from and why it is associative.Suppose we have a composition of L.T.'s [linear transformations]

$$\mathbf{V} \xrightarrow{\mathbf{T}} \mathbf{W} \xrightarrow{\mathbf{U}} \mathbf{Z},$$

and that we have chosen bases $\mathcal{B}$, $\mathcal{C}$ $\mathcal{D}$, so that we have matrix representations

$$[\mathbf{T}]_{\mathcal{B}}^{\mathcal{C}}, \quad [\mathbf{U}]_{\mathcal{C}}^{\mathcal{D}}.$$

Theorem

The matrix representation of the composition $\mathbf{U} \circ \mathbf{T}$ *is*

$$[\mathbf{U} \circ \mathbf{T}]_{\mathcal{B}}^{\mathcal{D}} = [\mathbf{U}]_{\mathcal{C}}^{\mathcal{D}} \circ [\mathbf{T}]_{\mathcal{B}}^{\mathcal{C}}.$$

To prove this we pick the bases $\mathcal{B} = v_1, \ldots, v_n$, $\mathcal{C} = u_1, \ldots, u_m$, $\mathcal{D} = w_1, \ldots, w_p$, and look for the coefficient c_{ji} of w_i in the expression of $(\mathbf{U} \circ \mathbf{T})(v_j)$:

$$\begin{aligned}
(\mathbf{U} \circ \mathbf{T})(v_j) &= \mathbf{U}(\mathbf{T}(v_j)) = \mathbf{U}(\sum_{k=1}^{m} a_{kj} u_k) \\
&= \sum_{k=1}^{m} a_{kj} \mathbf{U}(u_k) = \sum_{k=1}^{m} a_{kj} (\sum_{\ell=1}^{p} b_{\ell k} w_\ell) \\
&= \sum_{\ell=1}^{p} (\sum_{k=1}^{m} b_{\ell k} a_{kj}) w_\ell
\end{aligned}$$

This gives

$$c_{ij} = \sum_{k=1}^{m} b_{ik} a_{kj},$$

the usual row by column rule of multiplication.

There is a consequence that is tedious to verify directly, that the product of matrices is associative:

This follows from the tautology of the composition of functions

$$(\mathbf{A} \circ \mathbf{B}) \circ \mathbf{C} = \mathbf{A}(\mathbf{B}(\mathbf{C})) = \mathbf{A} \circ (\mathbf{B} \circ \mathbf{C})$$

and the theorem above.

Let us solve an exercise that usually gets a shaky argument. Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ matrices such that $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}$.

Claim: $\mathbf{B} \cdot \mathbf{A} = \mathbf{I}$. [The question arises because matrix multiplication is not commutative.] To argue we consider the L.T.s $\mathbf{L_A}$ and $\mathbf{L_B}$ associated to $\mathbf{A}$ and $\mathbf{B}$.

$\mathbf{A} \cdot \mathbf{B} = \mathbf{I}$ implies that

$$\mathbf{L_A} \circ \mathbf{L_B} = \mathbf{I},$$

from which it follows that $\mathbf{L_B}$ is one-one, and therefore it is invertible, so

$$\mathbf{L_A} \circ \mathbf{L_B} = \mathbf{L_B} \circ \mathbf{L_A} = \mathbf{I}.$$

Thus $\mathbf{B} \cdot \mathbf{A} = \mathbf{I}$.

Outline

HomeQuiz + Quiz #3

For next week there will be two parts of the HomeQuiz: (1) a typical class quiz based in the following Homework:

1. Section 1.5: 2(g), 8, 10, 19
2. Section 1.6: 4, 14, 30, 34.
3. Section 2.1: 19, 25.
4. Section 2.2: 11.

HomeQuiz + Quiz #3, Cont'd

A typical HomeQuiz for the following Problems:

1. Let $\mathbf{M}_2(\mathbf{F})$ be the vector space of all 2-by-2 matrices over the field $\mathbf{F}$. Fix a matrix, say, $\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}$ and define the function

$$\mathbf{B} \rightarrow \mathbf{T}(\mathbf{B}) = \mathbf{AB}.$$

   We saw that $\mathbf{T}$ is a linear transformation. Find a matrix representation of $\mathbf{T}$.

2. If $\mathbf{F} = \mathbb{Z}_2$, find bases of the nullspace and range of $\mathbf{T}$.

3. Find a real polynomial $p(x)$, of degree at most 3, passing through the points of coordinates $(0, 1), (1, 3), (3, -1), (4, 0)$.

Outline

Invertible linear transformations

Let

$$\mathbf{T} : \mathbf{V} \to \mathbf{W}$$

be a L.T. that is one-one and onto. This means that for any $w \in \mathbf{W}$ there is a unique $v \in \mathbf{V}$ such that $\mathbf{T}(v) = w$. This gives rise to a function

$$\mathbf{U} : \mathbf{W} \to \mathbf{V}, \quad \mathbf{U}(w) = v \quad \text{iff} \quad \mathbf{T}(v) = w.$$

U is the inverse function of **T**:

$$\mathbf{U} \circ \mathbf{T} = \mathbf{I}_{\mathbf{V}} \quad \text{the identity of } \mathbf{V}.$$

One also checks

$$\mathbf{T} \circ \mathbf{U} = \mathbf{I}_{\mathbf{W}} \quad \text{the identity of } \mathbf{W}.$$

Notation: $\mathbf{U} = \mathbf{T}^{-1}$.

Proposition

If **T** *is a L.T. then* **U** *is also a L.T.*

Proof.

Let $w_1, w_2 \in \mathbf{W}$. Pick $v_1, v_2 \in \mathbf{V}$ so that $\mathbf{T}(v_1) = w_1$ and $\mathbf{T}(v_2) = w_2$. Since **T** is a L.T.,

$$\mathbf{T}(v_1 + v_2) = \mathbf{T}(v_1) + \mathbf{T}(v_2) = w_1 + w_2.$$

By the definition of **U**,

$$\mathbf{U}(w_1 + w_2) = v_1 + v_2 = \mathbf{U}(w_1) + \mathbf{U}(w_2),$$

so **U** is additive. The scaling property is proved in the same way. □

If $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ is an invertible L.T., choosing bases $\mathcal{B}$ and $\mathcal{C}$ for the two spaces:

Proposition

The matrix representations of $\mathbf{T}$ *and* $\mathbf{T}^{-1}$ *are related as follows*

$$[\mathbf{T}^{-1}]_{\mathcal{C}}^{\mathcal{B}} = ([\mathbf{T}]_{\mathcal{B}}^{\mathcal{C}})^{-1}.$$

Proof.

This follows from the equalities

$$\mathbf{T}^{-1} \circ \mathbf{T} = \mathbf{I}_{\mathbf{V}}, \quad \mathbf{T} \circ \mathbf{T}^{-1} = \mathbf{I}_{\mathbf{W}}$$

and a previous result asserting that the matrix representation of a composition of two L.T. is the product of the matrices. □

If $\mathbf{T}$ is invertible, we also say that it is an isomorphism, and that $\mathbf{V}$ and $\mathbf{W}$ are isomorphic vector spaces. For this to happens it requires that $\dim \mathbf{V} = \dim \mathbf{W}$.

Example: Let $\mathbb{P}_4[x]$ be the space of polynomials of degree at most 4 with coefficients in the field $\mathbf{F}$. The mapping

$$\mathbf{T}(a_0 + a_1 x + \cdots + a_4 x^4) = (a_0, a_1, \ldots, a_4)$$

is an isomorphism between $\mathbb{P}_4[x]$ and $\mathbf{F}^5$.

Similarly, it is easy to define isomorphisms between $\mathbf{M}_n(\mathbf{F})$ and $\mathbf{F}^{n^2}$.

Examples

- A linear transformation $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ such that $\mathbf{T}^2 = 0$ obviously cannot be invertible. Note however that $\mathbf{I} - \mathbf{T}$ is always invertible:

$$(\mathbf{I} - \mathbf{T})(\mathbf{I} + \mathbf{T}) = \mathbf{I} - \mathbf{T}^2 = \mathbf{I}.$$

- Prove the same assertion if $\mathbf{T}^3 = 0$ [or any other power $\mathbf{T}^n = 0$].
- Let $\mathbf{V}$ be the vector space of all sequences $(a_1, a_2, a_3, \ldots)$. The functions right shift and left shift are L.T.

$$\begin{aligned} \mathbf{r}(a_1, a_2, a_3, \ldots) &= (0, a_1, a_2, \ldots) \\ \mathbf{s}(a_1, a_2, a_3, \ldots) &= (a_2, a_3, a_4 \ldots) \end{aligned}$$

$\mathbf{r}$ is one-one but not an isomorphism, $\mathbf{s}$ is onto but not an isomorphism.

Exercise 5: Let $\mathbf{A}$ be a fixed $n \times n$ of rank r. Define the mapping $\mathbf{T} : \mathbf{M}_n(\mathbf{F}) \to \mathbf{M}_n(\mathbf{F})$ by

$$\mathbf{B} \mapsto \mathbf{AB}.$$

Show that $\mathbf{T}$ is a linear transformation of rank $r \cdot n$.

Answer: Consider only the case $n = 2$.

- Check $\mathbf{T}$ is a L.T.:
  $\mathbf{T}(\mathbf{B}_1 + \mathbf{B}_2) = \mathbf{A}(\mathbf{B}_1 + \mathbf{B}_2) = \mathbf{AB}_1 + \mathbf{AB}_2 = \mathbf{T}(\mathbf{B}_1) + \mathbf{T}(\mathbf{B}_2)$ and
  $\mathbf{T}(c\mathbf{B}) = \mathbf{A}c\mathbf{B} = c(\mathbf{AB}) = c\mathbf{T}(\mathbf{B})$
- $(n = 2)$ If rank $\mathbf{A} = 0$, $\mathbf{A} = O$, then $\mathbf{T}$ is also the null mapping (so $\mathbf{T}$ has rank 0.
- If rank $\mathbf{A} = 2$, $\mathbf{A}$ is invertible and have for any matrix $\mathbf{B}$, $\mathbf{T}(\mathbf{A}^{-1}\mathbf{B}) = \mathbf{B}$, so $\mathbf{T}$ is an onto mapping of $\mathbf{M}_2(\mathbf{F})$, that is has rank 2^2.
- If $\mathbf{A}$ has rank 1, it has a form like $\begin{bmatrix} 1 & a \\ 0 & 0 \end{bmatrix}$. **Reader: Your turn to check rank T is 2**

Exercise 6: Show that there is no square nonzero real matrix $\mathbf{A}$ such that

$$\mathbf{A}^t = r\mathbf{A}, \quad r \neq \pm 1.$$

Answer: If $\mathbf{A}^t = r\mathbf{A}$, transposing this we get

$$\mathbf{A} = (\mathbf{A}^t)^t = (r\mathbf{A})^t = r\mathbf{A}^t = r^2\mathbf{A}$$

Thus $(1 - r^2)\mathbf{A} = O$, and therefore $r = \pm 1$ or $\mathbf{A} = O$.

Exercise 6a: Section 2.3: #15, #19

Outline

Quickly: Changing coordinates permit the solution of many problems. Here are two:

- To evaluate $\int_0^1 te^{t^2}dt$, one sets $y = t^2$ and the problem becomes

$$\int_0^1 te^{t^2}dt = \int_0^1 \frac{1}{2}e^y dy = \frac{1}{2}(e-1).$$

- What is the graph of the equation $2x^2 + 6xy + 10y^2 = 100$? The solution requires a change of point-of-view–which a change of coordinates will bring.

Example

- Let $\mathbf{T} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ be a L.T. of $\mathbb{R}^2$ in terms of the standard basis, $\{i, j\}$.
- The vectors $v_1 = i$ and $v_2 = i + j$ form another basis of $\mathbb{R}^2$. Note that $\mathbf{T}(v_1) = v_2$, and $\mathbf{T}(v_2) = (i + j) + (i - j) = 2i = 2v_1$.
- The matrix representation of $\mathbf{T}$ is $\begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix}$.

The change of coordinates issue we will discuss is the following: Let $v \in \mathbf{V}$ be a vector of a V.S. If two bases $\mathcal{A} = v_1, \ldots, v_n$ and $\mathcal{B} = u_1, \ldots, u_n$ are picked in $\mathbf{V}$, the vector has two representations

$$[v]_{\mathcal{A}} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad [v]_{\mathcal{B}} = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}$$

Question: How are the x_i related to the x_i'? The answer will depend on how the v_i and u_i relate.

$$[v]_{\mathcal{A}} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{P}[v]_{\mathcal{B}} = \mathbf{Q} \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}$$

Change of bases formula

We have

$$v = \sum_{j=1}^{n} x_j v_j = \sum_{j=1}^{n} x'_j u_j.$$

We start from

$$v_j = \sum_{i=1}^{n} p_{ij} u_i, \quad u_j = \sum_{i=1}^{n} q_{ij} v_i$$

Note the two [basis changing] matrices

$$\mathbf{P} = [p_{ij}], \quad \mathbf{Q} = [q_{ij}]$$

If we replace $v_j = \sum_{i=1}^{n} p_{ij} u_i$ in

$$v = \sum_{j=1}^{n} x_j v_j = \sum_{j=1}^{n} x_j' u_j.$$

we get

$$v = \sum_{j=1}^{n} x_j (\sum_{i=1}^{n} p_{ij} u_i) = \sum_{i=1}^{n} (\sum_{j=1}^{n} p_{ij} x_j) u_i = \sum_{i=1}^{n} x_i' u_i.$$

$$x_i' = \sum_{j=1}^{n} p_{ij} x_j,$$

the desired formulas.

In matrix notation:

$$\begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix} = \mathbf{P} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{Q} \cdot \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}$$

Note

$$\mathbf{P} \cdot \mathbf{Q} = \mathbf{I}$$

Max and Min of functions of several variables

The need for change of variables occur in the determination of local maxima and minima of functions of several variables. Recall that the function $\mathbf{f}(x, y)$ has a local maximum at (a, b) if

$$\mathbf{f}(a, b) \geq \mathbf{f}(x, y),$$

for (x, y) near (a, b). If $\mathbf{f}$ has derivatives, this first requires

$$\frac{\partial \mathbf{f}}{\partial x}(a, b) = \frac{\partial \mathbf{f}}{\partial y}(a, b) = 0$$

What else? We expand $\mathbf{f}$ around (a, b):

$$\begin{aligned} \mathbf{f}(x, y) &= \mathbf{f}(a, b) + (\underbrace{\mathbf{f}_x(a, b)(x - a) + \mathbf{f}_y(a, b)(y - b) = 0}) \\ &+ \underline{1/2(\mathbf{f}_{xx}(a, b)(x - a)^2 + 2\mathbf{f}_{xy}(x - a)(y - b) + \mathbf{f}_{yy}(y - b)^2)} + \end{aligned}$$

Whether (a, b) is a local maximum will depend on whether the term

$$\underline{1/2(\mathbf{f}_{xx}(a,b)(x-a)^2 + 2\mathbf{f}_{xy}(x-a)(y-b) + \mathbf{f}_{yy}(y-b)^2)}$$

is always non-positive near (a, b).
Hard to guess when a polynomial

$$Ax^2 + Bxy + Cy^2$$

is always negative near the origin, UNLESS $B = 0$ when the condition is $A, C \leq 0$. It involves the examination of

$$\begin{bmatrix} \mathbf{f}_{xx} & \mathbf{f}_{xy} \\ \mathbf{f}_{yx} & \mathbf{f}_{yy} \end{bmatrix}$$

Just imagine the size of the problem in 5 or 10 variables! Fortunately, Linear Algebra comes to the rescue: it involves a certain calculation with the matrix of second order derivatives. In the case of 3 variables,

$$\begin{bmatrix} \mathbf{f}_{xx} & \mathbf{f}_{xy} & \mathbf{f}_{xz} \\ \mathbf{f}_{yx} & \mathbf{f}_{yy} & \mathbf{f}_{yz} \\ \mathbf{f}_{zx} & \mathbf{f}_{zy} & \mathbf{f}_{zz} \end{bmatrix}$$

This is so important that we will have to return to the topic for a serious treatment.

Identifying conics

To identify the graph of the equation $2x^2 + 6xy + 10y^2 = 100$, we do the change of coordinates

$$\begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{P} \cdot \begin{bmatrix} x' \\ y' \end{bmatrix} = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix}$$

In the new coordinates the equation is

$$(x')^2 + 11(y')^2 = 100,$$

an ellipse.

Question: How did we find **P**?

Proposition

Suppose $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ *is a L.T. For each basis* $\mathcal{A} = v_1, \ldots, v_n$ *we have a matrix representation* $[\mathbf{T}]_{\mathcal{A}} = \mathbf{A}$. *Given another basis* $\mathcal{B} = u_1, \ldots, u_n$ *we have another matrix representation* $[\mathbf{T}]_{\mathcal{B}} = \mathbf{B}$. *Then*

$$\mathbf{A} = \mathbf{Q}^{-1}\mathbf{B}\mathbf{Q} = \mathbf{P}\mathbf{B}\mathbf{Q} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}.$$

Proof.

The matrices $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ appear from the expressions $\mathbf{T}(v_j) = \sum a_{ij} v_i$ and $\mathbf{T}(u_j) = \sum b_{ij} u_i$ and the relationship from:

$$\begin{aligned} \mathbf{T}(v_j) &= \mathbf{T}(\sum_{r=1}^{n} p_{rj} u_r) = \sum_{r=1}^{n} p_{rj} \mathbf{T}(u_r) \\ &= \sum_{r=1}^{n} \sum_{k=1}^{n} p_{rj} b_{kr} u_k = \sum_{r=1}^{n} \sum_{k=1}^{n} \sum_{\ell=1}^{n} p_{rj} b_{kr} q_{\ell k} v_\ell \end{aligned}$$

□

Observations

- Two $n \times n$ matrices are said to be similar if there is an invertible matrix $\mathbf{S}$ such that $\mathbf{A} = \mathbf{S}\mathbf{B}\mathbf{S}^{-1}$. [This is an equivalence relation.]
- This fact can be interpreted by saying that $\mathbf{A}$ and $\mathbf{B}$ are matrix representations of the same L.A. $\mathbf{L_A}$ but with respect to different bases.
- It is very important to find matrix representations which are simple. [Recall: diagonalization.]