Summary of week 2 (lectures 4, 5 and 6)

Lecture 4 was concerned with the concept of *linearity*. This material appears §3a of the text *Vector Space Theory* (referred to as [VST] below).

Definition Let V and W be vector spaces over the same field F. A function $f\colon V \to W$ is said to be *linear* if

$$f(\underset{\sim}{x} + \underset{\sim}{y}) = f(\underset{\sim}{x}) + f(\underset{\sim}{y})$$

and

$$f(\lambda \underset{\sim}{x}) = \lambda f(\underset{\sim}{x})$$

for all $\underset{\sim}{x}$, $\underset{\sim}{y} \in V$ and $\lambda \in F$.

Examples

1) Let $V = \mathbb{R}^3$ (the set of all 3-component column vectors whose entries are real numbers) and $W = \mathbb{R}^2$. The V and W are both vector spaces over $\mathbb{R}$. Let $f\colon V \to W$ be defined by

$$f\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} a+b+c \\ 3a+2c \end{pmatrix}.$$

It is left to the reader to verify that f is linear. Note that the definition of f could alternatively be written as

$$f\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 3 & 0 & 2 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix},$$

and the fact that f is linear is thus a consequence of well known properties of matrix multiplication: see (ii) and (iv) on p.21 of [VST]. These properties were covered in 1st year.

2) Let $\mathscr{D}$ be the set of all differentiable functions from $\mathbb{R}$ to $\mathbb{R}$, and $\mathscr{F}$ be the set of all functions from $\mathbb{R}$ to $\mathbb{R}$. Then $\mathscr{D}$ and $\mathscr{F}$ are vector spaces over $\mathbb{R}$. Define a function $T\colon \mathscr{D} \to \mathscr{F}$ as follows: if $f \in \mathscr{D}$ then $Tf \in \mathscr{F}$ is given by

$$(Tf)(t) = f'(t) + t^3 f(t) \qquad \text{for all } t \in \mathbb{R}.$$

Then T is linear. This is proved by showing that $T(f+g) = Tf + Tg$ and $T(\lambda f) = \lambda(Tf)$ for all f, $g \in \mathscr{D}$ and $\lambda \in \mathbb{R}$.
Recall that addition and scalar multiplication for functions are defined by the formulas

$$(f+g)(t) = f(t) + g(t)$$

and

$$(\lambda f)(t) = \lambda(f(t)).$$

It is a theorem of calculus that if real-valued functions f and g are differentiable at some point $t \in \mathbb{R}$ then so is $f+g$, and $\frac{d}{dt}(f(t)+g(t) = \frac{d}{dt}f(t)+\frac{d}{dt}g(t)$. Thus if $f,\, g \in \mathscr{D}$ then, for all $t \in \mathbb{R}$,

$$\begin{aligned}(T(f+g))(t) &= (f+g)'(t) + t^3((f+g)(t)) \quad \text{(by the definition of } T)\\ &= f'(t) + g'(t) + t^3(f(t)+g(t))\\ &= (f'(t) + t^3 f(t)) + (g'(t) + t^3 g(t))\\ &= (Tf)(t) + (Tg)(t)\\ &= (Tf+Tg)(t).\end{aligned}$$

So the functions $T(f+g)$ and $Tf+Tg$ take the same value at all $t \in \mathbb{R}$, and so $T(f+g) = Tf+Tg$. The proof that $T(\lambda f) = \lambda(Tf)$ is similar. See also the example on p.51 of [VST].

3) Let $\mathscr{C}$ be the set of all continuous functions $\mathbb{R} \to \mathbb{R}$, and define $S\colon \mathscr{C} \to \mathscr{F}$ by

$$(Sf)(t) = \int_0^t f(u)\, du.$$

Then S is linear; this is simply a statement of the following standard theorems of calculus:

$$\int_0^t f(u) + g(u)\, du = \int_0^t f(u)\, du + \int_0^t g(u)\, du$$

and

$$\int_0^t \lambda f(u)\, du = \lambda \int_0^t f(u)\, du$$

for all continuous functions f and g and all $\lambda \in \mathbb{R}$.

The importance of vector spaces derives from the fact that many problems that arise naturally in science or engineering involve linear functions, and vector space theory provides the framework to analyse linear functions. Linear equations of various kinds—equations of the form $T(x) = c$, where T is a linear function—are particularly common. Here are some examples of linear equations.

1) Find all x, y and z such that

$$\begin{pmatrix} 1 & 1 & 1 \\ 3 & 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

(You should know from 1st year how to solve systems like this.)

2) Find all differentiable functions $f\colon \mathbb{R} \to \mathbb{R}$ satisfying

$$f'(t) + t^3 f(t) = e^{-t^3}.$$

(This is an example of a linear differential equation.)

3) Find $f\colon \{\, x \in \mathbb{R} \mid x > 0 \,\} \to \mathbb{R}$ satisfying

$$\int_t^\infty f(u)\, du = \tfrac{1}{t}.$$

(Solving this simply amounts to differentiating $1/t$; nevertheless, it is a linear equation.)

The above examples, and all other linear equations, share the following important property, sometimes called the *principle of superposition*: if x_1 and x_2 are solutions of the equations $T(x) = c_1$ and $T(x) = c_2$ respectively, then $x_1 + x_2$ is a solution of $T(x) = c_1 + c_2$. The following theorem is essentially a statement of this principle.

Theorem. *Let $T\colon V\colon W$ be a linear function, where V and W are vector spaces over the same field F, and let $\underset{\sim}{a} \in W$ be fixed. Let $\underset{\sim}{0}_W$ denote the zero element of W, and set*

$$H = \{\, \underset{\sim}{x} \in V \mid T(\underset{\sim}{x}) = \underset{\sim}{0}_W \,\}$$

and

$$S = \{\, \underset{\sim}{x} \in V \mid T(\underset{\sim}{x}) = \underset{\sim}{a} \,\}.$$

If $\underset{\sim}{x}_0$ is some fixed element of S then

$$S = \{\, \underset{\sim}{x} + \underset{\sim}{z} \mid \underset{\sim}{z} \in H \,\}.$$

Alternatively put, if $\underset{\sim}{x}_0$ is a solution of the linear equation $T(\underset{\sim}{x}) = \underset{\sim}{a}$ then the general solution of this equation consists of all elements of the form $\underset{\sim}{x} = \underset{\sim}{x}_0 + \underset{\sim}{z}$, where $\underset{\sim}{z}$ is a solution of the "homogeneous" equation $T(\underset{\sim}{x}) = \underset{\sim}{0}_W$.

A proof of the above theorem was given in Lecture 5. The argument is essentially that given in Example #1 on p.51 of [VST]. To see that the argument does apply for any linear function T from one vector space to another, it is necessary to prove some elementary consequences of the vector space axioms, such as Propositions 3.4, 3.5, 3.6 and 3.7 of [VST]. Student should read through the proofs of these.

Lecture 5 also dealt with Theorem 2.2 of [VST], and multiplication of partitioned matrices (pp. 22, 23 of [VST]). The latter was illustrated by considering $2n \times 2n$ matrices of the form

$$\begin{pmatrix} I & A \\ 0 & I \end{pmatrix}$$

where here I and 0 denote (respectively) the $n \times n$ identity and zero matrices, and A can be any $n \times n$ matrix. It can be verified that

$$\begin{pmatrix} I & A \\ 0 & I \end{pmatrix} \begin{pmatrix} I & B \\ 0 & I \end{pmatrix} = \begin{pmatrix} I & A+B \\ 0 & I \end{pmatrix} \qquad (*)$$

for all A and B. So, in particular,

$$\begin{pmatrix} I & A \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 8 \\ 0 & I \end{pmatrix} = \begin{pmatrix} I & B \\ 0 & I \end{pmatrix} \begin{pmatrix} I & A \\ 0 & I \end{pmatrix}.$$

Furthermore, if we put $B = -A$ then the right hand side of $(*)$ is just the $2n \times 2n$ identity matrix, and so

$$\begin{pmatrix} I & A \\ 0 & I \end{pmatrix}^{-1} = \begin{pmatrix} I & -A \\ 0 & I \end{pmatrix}.$$

For example,

$$\begin{pmatrix} 1 & 0 & 2 & 5 \\ 0 & 1 & -4 & 3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & -2 & -5 \\ 0 & 1 & 4 & -3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Lecture 5 also included a brief reminder of the properties of determinants, and dealt with in 1st year. The relevant properties are stated on pp. 36, 37 and in Theorem 8.13 on p. 181 of [VST]. We shall study determinants in more detail, including a revision of the 1st year material, later in the semester.

Lecture 6 commenced with an account of elementary matrices and elementary row operations. If m is a positive integer and r, $s \in \{1, 2, \ldots, m\}$ then we define E_{rs} to be the $m \times m$ matrix whose (i, j)-entry is given by

$$(E_{rs})_{ij} = \begin{cases} \delta_{ij} & \text{if } i \neq r \text{ and } j \neq s \\ \delta_{sj} & \text{if } i = r \\ \delta_{rj} & \text{if } i = s \end{cases}$$

for all i, $j \in \{1, 2, \ldots, m\}$. Here δ_{ij} is the *Kronecker delta*, defined by

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

Thus δ_{ij} is the (i, j)-entry of the identity matrix I. The formula for $(E_{rs})_{ij}$ says that the i-th row of the matrix E_{rs} is the same as the i-th row of the identity matrix if $i \neq r$ and $i \neq s$, while the r-th row of E_{rs} is the s-th row of I and the s-th row of E_{rs} is the r-th row of I.

The matrices E_{rs} comprise one of three types of elementary matrices. The others are given as follows: if $r \in \{1, 2, \ldots, m\}$ and λ is a nonzero scalar then $E_r^{(\lambda)}$ is the matrix given by

$$(E_r^{(\lambda)})_{ij} = \begin{cases} \delta_{ij} & \text{if } i \neq r \\ \lambda\delta_{rj} & \text{if } i = r \end{cases}$$

and if r, $s \in \{1, 2, \ldots, m\}$ and λ is any scalar then $E_{rs}^{(\lambda)}$ is given by

$$(E_{rs}^{(\lambda)})_{ij} = \begin{cases} \delta_{ij} & \text{if } i \neq s \\ \delta_{sj} + \lambda\delta_{rj} & \text{if } i = s. \end{cases}$$

It can be seen that $E_r^{(\lambda)}$ is obtained by multiplying the r-th row of I by λ, while $E_{rs}^{(\lambda)}$ is obtained by adding λ times the r-th row of I to the s-th row.

Corresponding to the three kinds of elemntary matrices there are three kinds of *elementary row operations*. These can be regarded as functions defined on the set of all $m \times n$ matrices (where m and n are some fixed positive integers). Specifically, define ρ_{rs}, $\rho_r^{(\lambda)}$ and $\rho_{rs}^{(\lambda)}$ from $\mathrm{Mat}(m \times n, F)$ to $\mathrm{Mat}(m \times n, F)$ by

$$\begin{aligned} \rho_{rs}(A) &= E_{rs}A \\ \rho_r^{(\lambda)}(A) &= E_r^{(\lambda)}A \\ \rho_{rs}^{(\lambda)}(A) &= E_{rs}^{(\lambda)}A \end{aligned}$$

for all $A \in \mathrm{Mat}(m \times n, F)$. (Here, as always, F is the field of scalars; it remains fixed in any given context.)

We have the following theorem:

Theorem. *In the above notation, $\rho_{rs}(A)$ is the result of swapping the r-th and s-th rows of A, while $\rho_r^{(\lambda)}(A)$ is the result of multipying the r-th row of A by λ, and $\rho_{rs}^{(\lambda)}(A)$ is the result of adding λ times the r-th row of A to the s-th row.*

The last part of this was proved in the lecture. Here is a proof of the second part. Let $r \in \{1, 2, \ldots, m\}$ and $0 \neq \lambda \in F$. Then for all $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$,

$$\begin{aligned} (\rho_r^{(\lambda)}(A))_{ij} &= (E_r^{(\lambda)}A)_{ij} \\ &= \sum_{k=1}^{m} (E_r^{(\lambda)})_{ik} A_{kj} \\ &= \begin{cases} \sum_{k=1}^{m} \delta_{ik} A_{kj} & \text{if } i \neq r \\ \sum_{k=1}^{m} \lambda\delta_{rk} A_{kj} & \text{if } i = r \end{cases} \\ &= \begin{cases} A_{ik} & \text{if } i \neq r \\ \lambda A_{rk} & \text{if } i = r. \end{cases} \end{aligned}$$

This shows that the i-th row of $\rho_r^{(\lambda)}(A)$ is the same as the i-th row of A if $i \neq r$, while the r-th row of $\rho_r^{(\lambda)}(A)$ is λ times the r-th row of A, as required.

Note that in the course of this proof we used the following "substitution property" of the Kronecker delta: $\sum_j \delta_{ij} a_j = a_i$. The truth of this is easily seen, since the only nonzero term in the sum occurs when j (the index of summation) is equal to i. (It is assumed that i lies in the range of values that j runs through.)

Elementary matrices and elementary row operations are discussed on pp. 32, 33 of [VST]. However, [VST] takes a slightly different approach from that taken in lectures, in that the statement of Theorem 2.6 of [VST] was taken in lectures as the definition of the elementary row operations, and consequently Definition 2.3 of [VST] became a theorem.

The discussion of elementary matrices was followed by the proof of the important fact that $\det(AB) = (\det A)(\det B)$ for all $m \times m$ matrices A and B. Again, this is revision of 1st year material. The proof can also be found on pp. 184–186 of [VST] (Proposition 8.14 through to Theorem 8.20).

Lecture 6 concluded with a revision of eigenvalues and eigenvectors (see Definition 2.11 of [VST]). In particular, diagonalization of a square matrix A was discussed: given A, find (if possible) an invertible matrix T such that

$$T^{-1}AT = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

for some scalars $\lambda_1, \lambda_2, \dots, \lambda_n$. Writing D for the diagonal matrix on the right hand side, it is easily seen that the above equation is equivalent to $AT = TD$ (given that T is invertible). Now if we write $\underset{\sim}{c}_i$ for the i-th column of the matrix T then the i-column of AT is $A\underset{\sim}{c}_i$. On the other hand, the i-th column of TD is

$$T \begin{pmatrix} 0 \\ \vdots \\ \lambda_i \\ \vdots \\ 0 \end{pmatrix} = \left(\underset{\sim}{c}_1 \middle| \dots \middle| \underset{\sim}{c}_i \middle| \dots \middle| \underset{\sim}{c}_n \right) \begin{pmatrix} 0 \\ \vdots \\ \lambda_i \\ \vdots \\ 0 \end{pmatrix} = \underset{\sim}{c}_1 0 + \dots + \underset{\sim}{c}_i \lambda_i + \dots + \underset{\sim}{c}_n 0 = \lambda_i \underset{\sim}{c}_i.$$

Thus we conclude that the matrix equation above holds if and only if $A\underset{\sim}{c}_i = \lambda_i \underset{\sim}{c}_i$ for all i. It follows that A is diagonalizable if and only if there is an invertible matrix T whose columns are all eigenvectors of A.