MATH 2080 Further Linear Algebra

Jonathan R. Partington, University of Leeds, School of Mathematics

December 8, 2010

LECTURE 1

Books:
S. Lipschutz – Schaum's outline of linear algebra
S.I. Grossman – Elementary linear algebra

1 Vector spaces and subspaces

Vector spaces have two built-in concepts.
1. Vectors – can be added or subtracted. Usually written $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, etc.
2. Scalars – can be added, subtracted, multiplied or divided (not by 0). Usually written a, b, c, etc.

Key example
$\mathbb{R}^n$, space of n-tuples of real numbers, $u = (u_1, \ldots, u_n)$.
If $\mathbf{u} = (u_1, \ldots, u_n)$ and $\mathbf{v} = (v_1, \ldots, v_n)$, then $\mathbf{u} + \mathbf{v} = (u_1 + v_1, \ldots, u_n + v_n)$.
Also if $a \in \mathbb{R}$, then $a\mathbf{u} = (au_1, \ldots, au_n)$.

1.1 *The obvious properties of the vector space* $\mathbb{R}^n$

(1) Vector addition satisfies:

- For all $\mathbf{u}$, $\mathbf{v}$, we have $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$, (commutative rule).
- For all $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, we have $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$, (associative rule).
- There is a *zero vector* $\mathbf{0}$ with $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$ for all $\mathbf{u}$.
- For all $\mathbf{u}$ there is an *inverse vector* $-\mathbf{u}$ with $\mathbf{u} + (-\mathbf{u}) = \mathbf{0} = (-\mathbf{u}) + \mathbf{u}$.

(2) Scalar multiplication satisfies:

- $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$ and

- $(a+b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$, (these are the distributive laws).
- $(ab)\mathbf{u} = a(b\mathbf{u})$, (associativity of scalar multiplication).
- $1\mathbf{u} = \mathbf{u}$, (identity property).

Now we look for other objects with the same properties.

Note our vectors were in $\mathbb{R}^n$, our scalars in $\mathbb{R}$. Instead of $\mathbb{R}$ we can use any set in which we have all the usual rules of arithmetic (a *field*).

Examples

$\mathbb{Q}$ – rational numbers (fractions) a/b, where a, b are integers and $b \neq 0$.
$\mathbb{C}$ – complex numbers.
A new one: $\mathbb{F}_2$ – the field of two elements, denoted 0 and 1, with usual rules of additions and multiplication except that $1 + 1 = 0$ (i.e., addition mod 2). So (-1) is the same as 1. This is used in coding theory, geometry, algebra, computer science, etc.

1.2 *Definition of a vector space*

A vector space V over a field F (which in this module can be $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ or $\mathbb{F}_2$) is a set V on which operations of vector addition $\mathbf{u}+\mathbf{v} \in V$ and scalar multiplication $a\mathbf{u} \in V$ have been defined, satisfying the eight rules given in (1.1).

Examples

(a) $V = F^n$, where $F = \mathbb{Q}, \mathbb{R}, \mathbb{C}$ or $\mathbb{F}_2$.
(b) $V = M_{m,n}$, all $m \times n$ matrices with entries in F.
(c) $V = P_n$, polynomials of degree at most n, i.e., $p(t) = a_0 + a_1 t + \ldots + a_n t^n$, with $a_0, a_1, \ldots, a_n \in F$.
(d) $V = F^X$. Let X be any set; then F^X is the collection of functions from X into F.
Define $(f + g)(x) = f(x) + g(x)$ and $(af)(x) = af(x)$, for $f, g \in V$ and $a \in F$.

1.3 *Other properties of a vector space*

We can deduce the following from the axioms in (1.1):
$a\mathbf{0} = \mathbf{0}$, for $a \in F$ and $\mathbf{0} \in V$.
$0\mathbf{v} = \mathbf{0}$, for $0 \in F$ and $\mathbf{v} \in V$.
If $a\mathbf{v} = \mathbf{0}$ then either $a = 0$ or $\mathbf{v} = 0$.
$(-1)\mathbf{v} = -\mathbf{v}$, and in general $(-a)\mathbf{v} = -(a\mathbf{v})$, for $a \in F$ and $\mathbf{v} \in V$.

The proofs are mostly omitted, but are short. For example, $a\mathbf{0} = a(\mathbf{0}+\mathbf{0}) = a\mathbf{0}+a\mathbf{0}$. Add $-(a\mathbf{0})$ to both sides and we get $\mathbf{0} = a\mathbf{0}+a\mathbf{0}+(-a\mathbf{0}) = a\mathbf{0}+\mathbf{0} = a\mathbf{0}$.

LECTURE 2

Subspaces

1.4 *Definition*

Let V be a vector space over a field F and W a subset of V. Then W is a *subspace* if it satisfies:
(i) $\mathbf{0} \in W$.
(ii) For all $\mathbf{v}, \mathbf{w} \in W$ we have $\mathbf{v}+\mathbf{w} \in W$.
(iii) For all $a \in F$ and $\mathbf{w} \in W$ we have $a\mathbf{w} \in W$.

That is, W contains $\mathbf{0}$ and is closed under the vector space operations. It's easy to see that then W is also a vector space, i.e., satisfies the properties of (1.1). For example $-\mathbf{w} = (-1)\mathbf{w} \in W$ if $\mathbf{w} \in W$.

1.5 *Examples*

(i) Every vector space V has two trivial subspaces, namely $\{\mathbf{0}\}$ and V.
(ii) Take any $\mathbf{v} \in V$, not the zero vector. Then $\operatorname{span}\{\mathbf{v}\} = \{a\mathbf{v} : a \in F\}$ is a subspace.
For example, in $\mathbb{R}^2$ we get a line through the origin [DIAGRAM]. These are the only subspaces of $\mathbb{R}^2$ apart from the trivial ones.
(iii) In $\mathbb{R}^3$ we have the possibilities in (i) and (ii) above, but we also have planes through the origin, e.g.,

$$W = \{(x,y,z) \in \mathbb{R}^3 : x-2y+3z = 0\}.$$

The general solution is obtained by fixing y and z, and then x is uniquely determined, e.g., $z = a$, $y = b$ and $x = -3a+2b$. So

$$\begin{aligned} W &= \{(-3a+2b, b, a) : a, b \in \mathbb{R}\} \\ &= \{a(-3,0,1)+b(2,1,0) : a, b \in \mathbb{R}\}. \end{aligned}$$

So we can see W either as all vectors orthogonal to $(1,-2,3)$, or all "linear combinations" of $(-3,0,1)$ and $(2,1,0)$ (two parameters).

1.6 *Definition*

Given a set S of vectors in V, the smallest subspace of V containing S is written $W = \operatorname{span}(S)$ or $\operatorname{lin}(S)$, and called the *linear span* of S.

It consists of all linear combinations $a_1\mathbf{s}_1 + a_2\mathbf{s}_2 + \ldots + a_n\mathbf{s}_n$, where $a_1, \ldots, a_n \in F$ and $\mathbf{s}_1, \ldots, \mathbf{s}_n \in S$. It includes $\mathbf{0}$ the "empty combination".
Note that all these combinations must lie in any subspace containing S, and if we add linear combinations or multiply by scalars, we still get a combination. So this is the smallest subspace containing S.

Example

In $\mathbb{R}^2$ the smallest subspace containing $(1,1)$ and $(2,3)$ is $\mathbb{R}^2$ itself, as we can write any (x,y) as $a(1,1)+b(2,3)$, solving $a+2b=x$ and $a+3b=y$ (uniquely). Whereas, span$\{(1,1),(2,2)\}$ is just span$\{(1,1)\}$ again.

1.7 *Proposition*

Let V be a vector space over F, and let U and W be subspaces of V. Then $U \cap W$ is also a subspace of V.
Proof: (i) $\mathbf{0} \in U \cap W$, since $\mathbf{0} \in U$ and $\mathbf{0} \in W$.
(ii) If $\mathbf{u}, \mathbf{v} \in U \cap W$, then $\mathbf{u}+\mathbf{v} \in U$ and $\mathbf{u}+\mathbf{v} \in W$, since each of $\mathbf{u}$ and $\mathbf{v}$ are, so $\mathbf{u}+\mathbf{v} \in U \cap W$.
(iii) Similarly if $a \in F$ and $\mathbf{u} \in U \cap W$, then $a\mathbf{u} \in U$ and $a\mathbf{u} \in W$ so $a\mathbf{u} \in U \cap W$.
$\square$

However, $U \cup W$ doesn't need to be a subspace. For example, in $\mathbb{R}^2$, take
$U = \{(x,0) : x \in \mathbb{R}\}$ and $W = \{(0,y) : y \in \mathbb{R}\}$. [DIAGRAM]
Then $(1,0) \in U \cup W$ and $(0,1) \in U \cup W$, but their sum is $(1,1) \notin U \cup W$.

LECTURE 3

Sums of subspaces

1.8 *Definition*

Let V be a vector space over a field F and U, W subspaces of V. Then
$U + W = \{\mathbf{u}+\mathbf{w} : \mathbf{u} \in U, \mathbf{w} \in W\}$.

1.9 *Proposition*

$U+W$ is a subspace of V, and is the smallest subspace containing both U and W.

Proof: (i) $\mathbf{0} = \mathbf{0}+\mathbf{0} \in U+W$ as $\mathbf{0} \in U$ and $\mathbf{0} \in W$.
(ii) If $\mathbf{v}_1 = \mathbf{u}_1 + \mathbf{w}_1$ and $\mathbf{v}_2 = \mathbf{u}_2 + \mathbf{w}_2$ are in $U+W$, then

$$\mathbf{v}_1 + \mathbf{v}_2 = \underbrace{(\mathbf{u}_1+\mathbf{u}_2)}_{\in U} + \underbrace{(\mathbf{w}_1+\mathbf{w}_2)}_{\in W} \in U+W.$$

(iii) If $\mathbf{v} = \mathbf{u} + \mathbf{w} \in U + W$ and $a \in F$, then

$$a\mathbf{v} = \underbrace{a\mathbf{u}}_{\in U} + \underbrace{a\mathbf{w}}_{\in W} \in U + W.$$

Every $\mathbf{u} \in U$ can be written as

$$\mathbf{u} = \underbrace{\mathbf{u}}_{\in U} + \underbrace{\mathbf{0}}_{\in W} \in U + W.$$

so $\mathbf{u} \in U + W$ and $U + W$ contains U (and W similarly). But any subspace containing U and W contains all vectors $\mathbf{u} + \mathbf{w}$, so $U + W$ is the smallest one.

□

Example
In $\mathbb{R}^3$ let $U = \{a(1,0,0) : a \in \mathbb{R}\}$, $W = \{b(0,1,0) : b \in \mathbb{R}\}$, and
$T = \{(c, d, -c) : c, d \in \mathbb{R}\}$.

Now $U + W = \{a(1,0,0) + b(0,1,0) : a, b \in \mathbb{R}\} = \{(a, b, 0) : a, b \in \mathbb{R}\}$.

Whereas $U + T = \mathbb{R}^3$, since, given $(x, y, z) \in \mathbb{R}^3$, we want to write
$(x, y, z) = \mathbf{u} + \mathbf{t} = (a, 0, 0) + (c, d, -c)$, i.e., to solve, $x = a + c$, $y = d$ and $z = -c$ for a, c and d. We can if $c = -z$, $d = y$ and $a = x + z$.

Also, $W + T = T$ since $W \subset T$, so any vector $\mathbf{w} + \mathbf{t}$ is already in T and we get nothing else.

1.10 *Definition*

In a vector space V with subspaces U and W, we say that $U + W$ is a *direct sum*, written $U \oplus W$, if $U \cap W = \{\mathbf{0}\}$.

In particular, $U \oplus W = V$ means $U + W = V$ and $U \cap W = \{\mathbf{0}\}$.

Examples
As above, $U \cap W = \{\mathbf{0}\}$, since if $(a, 0, 0) = (0, b, 0)$, then they are both $\mathbf{0}$.
So $U \oplus W = \{(a, b, 0) : a, b \in \mathbb{R}\}$.

$U \cap T = \{(0, 0, 0)\}$, as if $(a, 0, 0) = (c, d, -c)$ then $c = d = 0$. So $U \oplus T = \mathbb{R}^3$.

$W \cap T$ consists of all vectors $(0, b, 0) = (c, d, -c)$, for some b, c, d, which is all vectors $(0, b, 0)$, or W again.
So $W + T$ is not a direct sum, and the notation $W \oplus T$ is incorrect here.

1.11 *Proposition*

$V = U \oplus W$ if and only if for each $\mathbf{v} \in V$ there are **unique** $\mathbf{u} \in U$ and $\mathbf{w} \in W$ with $\mathbf{v} = \mathbf{u} + \mathbf{w}$.

Proof:
"⇒" The $\mathbf{u}$ and $\mathbf{w}$ are unique, since if $\mathbf{u}_1 + \mathbf{w}_1 = \mathbf{u}_2 + \mathbf{w}_2$ then

$$\underset{\in U}{\mathbf{u}_1 - \mathbf{u}_2} = \underset{\in W}{\mathbf{w}_2 - \mathbf{w}_1}$$

and, since $U \cap W = \{\mathbf{0}\}$, we have $\mathbf{u}_1 = \mathbf{u}_2$ and $\mathbf{w}_1 = \mathbf{w}_2$.

"⇐" If $\mathbf{v} \in U \cap W$, then

$$\mathbf{v} = \underset{\in U}{\mathbf{v}} + \underset{\in W}{\mathbf{0}} = \underset{\in U}{\mathbf{0}} + \underset{\in W}{\mathbf{v}}$$

and by uniqueness, $\mathbf{v} = \mathbf{0}$. So it's a direct sum.

□

In our example, since $U \oplus T = \mathbb{R}^3$, we can write any vector $\mathbf{v}$ in $\mathbb{R}^3$ uniquely as $\mathbf{v} = \mathbf{u} + \mathbf{t}$, with $\mathbf{u} \in U$ and $\mathbf{t} \in T$. For example, let's take $\mathbf{v} = (5, 6, 7)$. Then

$$(5, 6, 7) = (a, 0, 0) + (c, d, -c) = (a + c, d, -c)$$

gives $a = 12$, $d = 6$ and $c = -7$, i.e.,

$$(5, 6, 7) = (12, 0, 0) + (-7, 6, 7).$$

LECTURE 4

2 Linear dependence, spanning and bases

2.1 *Definition*

Let V be a vector space over a field F. Then a vector $\mathbf{v} \in V$ is a *linear combination* of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in V if we can write $\mathbf{v} = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n$ for some $a_1, \ldots, a_n \in F$.

2.2 *Definition*

A set if vectors $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is *linearly independent*, if the only solution to $a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n = \mathbf{0}$ is $a_1 = a_2 = \ldots = a_n = 0$.

This is the same as saying that we can't express any vector in S as a linear combination of the others.

2.3 *Examples:*

1) In $\mathbb{R}^3$, the vectors $\mathbf{v}_1 = (1, 0, 0)$, $\mathbf{v}_2 = (0, 1, 0)$ and $\mathbf{v}_3 = (0, 0, 1)$ are **independent**, since
$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 = (a_1, a_2, a_3) = \mathbf{0}$ only if $a_1 = a_2 = a_3 = 0$.

2) In $\mathbb{R}^3$, the vectors $\mathbf{v}_1 = (1, 0, 2)$, $\mathbf{v}_2 = (1, 1, 0)$ and $\mathbf{v}_3 = (-1, -2, 2)$ are linearly **dependent**, since $\mathbf{v}_1 - 2\mathbf{v}_2 - \mathbf{v}_3 = \mathbf{0}$. We can write any vector in terms of the others, e.g. $\mathbf{v}_3 = \mathbf{v}_1 - 2\mathbf{v}_2$.

2.4 *Definition*

A set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ *spans* V if every $\mathbf{v} \in V$ can be written as
$\mathbf{v} = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n$ for some $a_1, \ldots, a_n \in F$.

2.5 *Examples*

1) $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$ span $\mathbb{R}^3$.

2) See after (1.6). The set $\{(1, 1), (2, 3)\}$ spans $\mathbb{R}^2$ (the set of linear combinations is all of $\mathbb{R}^2$), whereas $\{(1, 1), (2, 2)\}$ **doesn't**.

2.6 *Definition*

If a set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ spans V and is linearly independent, then it is called a *basis* of V.

2.7 *Proposition*

$\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of V if and only if every $\mathbf{v} \in V$ can be written as a *unique* linear combination $v = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n$.

Proof:
Suppose that it is a basis. If there were two such ways of writing
$\mathbf{v} = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n = b_1\mathbf{v}_1 + \ldots + b_n\mathbf{v}_n$, then
$\mathbf{0} = (a_1 - b_1)\mathbf{v}_1 + \ldots + (a_n - b_n)\mathbf{v}_n$, and by linear independence we get
$a_1 - b_1 = 0, \quad \ldots, \quad a_n - b_n = 0$, which is uniqueness.

Conversely, if we always have uniqueness, we need to show the vectors are independent. But if $a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n = \mathbf{0}$, we know already that $0\mathbf{v}_1 + \ldots + 0\mathbf{v}_n = \mathbf{0}$, and so by uniqueness, $a_1 = \ldots = a_n = 0$, as required.

□

2.8 *Examples*

1) Clearly $\{(1,0,0),(0,1,0),(0,0,1)\}$ is a basis of $\mathbb{R}^3$.

2) Let's check whether $\{(1,1,1),(1,1,0),(1,0,0)\}$ is a basis of $\mathbb{R}^3$. We need to solve $x(1,1,1)+y(1,1,0)+z(1,0,0)=(a,b,c)$ for any given a,b,c.
That is

$$\begin{array}{ccccccc} x & + & y & + & z & = & a \\ x & + & y & & & = & b \\ x & & & & & = & c, \end{array}$$

with solution $x=c$, $y=b-c$, $z=a-b$ (solve from the bottom upwards). This is unique, so it's a basis.

3) Now try $\{(1,1,2),(1,2,0),(3,4,4)\}$ in $\mathbb{R}^3$. We solve
$x(1,1,2)+y(1,2,0)+z(3,4,4)=(a,b,c)$, i.e.,

$$\begin{array}{ccccccc} x & + & y & + & 3z & = & a \\ x & + & 2y & + & 4z & = & b \\ 2x & & & + & 4z & = & c, \end{array}$$

Row-reduce:

$$\begin{pmatrix} 1 & 1 & 3 & | & a \\ 1 & 2 & 4 & | & b \\ 2 & 0 & 4 & | & c \end{pmatrix} \overset{R2-R1,\ R3-2R1}{\longrightarrow} \begin{pmatrix} 1 & 1 & 3 & | & a \\ 0 & 1 & 1 & | & b-a \\ 0 & -2 & -2 & | & c-2a \end{pmatrix}$$

$$\overset{R3+2R2}{\longrightarrow} \begin{pmatrix} 1 & 1 & 3 & | & a \\ 0 & 1 & 1 & | & b-a \\ 0 & 0 & 0 & | & c+2b-4a \end{pmatrix},$$

which is equivalent to

$$\begin{array}{ccccccc} x & + & y & + & 3z & = & a \\ & & y & + & z & = & b-a \\ & & & & 0 & = & c+2b-4a, \end{array}$$

so we don't always get a solution – we only do if $c+2b-4a=0$, and the solution is not unique when it exists.
Indeed, $2(1,1,2)+(1,2,0)-(3,4,4)=\mathbf{0}$.

LECTURE 5

We are aiming to show that all bases of a vector space have the same number of elements.

2.9 *Exchange Lemma*

Let V be a vector space over a field F, and suppose that $\{\mathbf{u}_1, \ldots \mathbf{u}_n\}$ spans V. Let $\mathbf{v} \in V$ with $\mathbf{v} \neq \mathbf{0}$. Then we can replace $\mathbf{u}_j$ by $\mathbf{v}$ for some j with $1 \leq j \leq n$, so that the new set still spans V.

Proof: Since $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ spans, we can write $\mathbf{v} = a_1\mathbf{u}_1 + \ldots + a_n\mathbf{u}_n$, and since $\mathbf{v} \neq \mathbf{0}$ there is at least one non-zero a_j. Choose one. We have

$$\mathbf{u}_j = \frac{1}{a_j}(\mathbf{v} - (a_1\mathbf{u}_1 + \ldots + a_n\mathbf{u}_n) + a_j\mathbf{u}_j), \qquad (*)$$

which is a linear combination of $\mathbf{v}$ and all the $\mathbf{u}_1, \ldots, \mathbf{u}_n$ except $\mathbf{u}_j$.
Now any $\mathbf{w} = b_1\mathbf{u}_1 + \ldots + b_n\mathbf{u}_n$ can be written using (*) as a linear combination that uses $\mathbf{v}$ but not $\mathbf{u}_j$. So the new set spans.

□

2.10 *Theorem*

Let V be a vector space, let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be an independent set in V, and let $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be a spanning set. Then $n \geq k$ and we can delete k of the $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$, replacing them by $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$, so that the new set spans.

Proof: [Non-examinable: only a sketch given in lectures]
We'll apply (2.9) repeatedly. Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an independent set, none of the vectors are $\mathbf{0}$. So, after relabelling the spanning set if necessary, we can assume that $\{\mathbf{v}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ spans.

So $\mathbf{v}_2 = a_1\mathbf{v}_1 + a_2\mathbf{u}_2 + \ldots + a_n\mathbf{u}_n$ for some $a_1, \ldots, a_n$. We can't have $a_2 = \ldots = a_n = 0$ as then $\mathbf{v}_2 = a_1\mathbf{v}_1$, which contradicts independence. Without loss of generality, by relabelling, we can suppose $a_2 \neq 0$. Then exchange $\mathbf{u}_2$ for $\mathbf{v}_2$ to get $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{u}_3 \ldots, \mathbf{u}_n\}$ spanning.

Continue. Finally $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ spans and $k \leq n$.

□

2.11 *Example*

Take $V = \mathbb{R}^3$. Then $\mathbf{u}_1 = (1, 0, 0)$, $\mathbf{u}_2 = (0, 1, 0)$, $\mathbf{u}_3 = (0, 0, 1)$ span, and $\mathbf{v}_1 = (1, 1, 0)$ and $\mathbf{v}_2 = (1, 2, 0)$ are independent.

Now, $\mathbf{v}_1 = (1, 1, 0) = 1\mathbf{u}_1 + 1\mathbf{u}_2 + 0\mathbf{u}_3$, so we can replace either $\mathbf{u}_1$ or $\mathbf{u}_2$ by $\mathbf{v}_1$. Let's replace $\mathbf{u}_2$. Then $\{\mathbf{v}_1, \mathbf{u}_1, \mathbf{u}_3\}$ spans V.

So $\mathbf{v}_2 = (1, 2, 0) = a(1, 1, 0) + b(1, 0, 0) + c(0, 0, 1)$. Solving we get $a = 2, b = -1$ and $c = 0$. That is, $\mathbf{v}_2 = 2\mathbf{v}_1 - \mathbf{u}_1$. This means we can replace $\mathbf{u}_1$ by $\mathbf{v}_2$ and then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{u}_3\}$ spans V.

2.12 *Theorem*

Let V be a vector space and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be bases of V. Then $k = n$.

Proof: $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ are independent and $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ span, so $k \leq n$, by (2.10).
$\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ are independent and $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ span, so $n \leq k$, by (2.10).
So $n = k$.

□

2.13 *Definition*

A vector space V has *dimension* n, if it has a basis with exactly n elements (here $n \in \{1, 2, 3, \ldots\}$). It has *dimension 0*, if $V = \{\mathbf{0}\}$, only.
We call V *finite-dimensional* in this case, and write $\dim V = n$, where
$n \in \{0, 1, 2, \ldots\}$.

2.14 *Examples*

(i) F^n has dimension n over F, since it has basis
$\{(1, 0, \ldots, 0\}, (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)\}$, the *standard basis.*

(ii) P_n (polynomials of degree $\leq n$) has basis $\{1, t, t^2, \ldots, t^n\}$, so has dimension $n + 1$.

(iii) $\mathbb{C}^n$ is a vector space over $\mathbb{R}$ with dimension $2n$. A basis is
$\{(1, 0, \ldots, 0\}, (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1),$
$(i, 0, \ldots, 0\}, (0, i, 0, \ldots, 0), \ldots, (0, \ldots, 0, i)\}$.
This is not a basis when we use $\mathbb{C}$ as our scalars, since it is then no longer independent.

2.15 *Theorem*

Let V be a vector space of dimension n. Then any independent set has $\leq n$ elements, and, if it has exactly n, then it is a basis.
Any spanning set has $\geq n$ elements, and if it has exactly n, then it is a basis.

LECTURE 6

Proof: Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of V (i.e., spanning and independent). Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be an independent set. By (2.10), we have $k \leq n$. We can replace k of the $\mathbf{v}$'s by $\mathbf{u}$'s, so that it spans. So if $k = n$, then $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ spans and hence it is a basis.

Now let $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ be a spanning set. Then (2.10) tells us that $m \geq n$. Suppose now that $m = n$. If $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ is not independent, then $a_1\mathbf{w}_1 + \ldots + a_m\mathbf{w}_m = \mathbf{0}$, where at least one $a_i \neq 0$. But then $\mathbf{w}_i$ is a linear combination of the others, so we can delete it and the set of $n-1$ remaining vectors still spans. This is a contradiction, since a spanning set must always be at least as big as any independent set, by (2.10).

$\square$

2.16 *Examples*

In $\mathbb{R}^3$, the set $\{(1,2,3),(0,1,0)\}$ is independent, but can't be a basis, as not enough elements for it to span.

Also, $\{(1,2,3),(4,5,6),(0,1,0),(0,0,1)\}$ spans, but can't be a basis as there are too many elements and it is not independent.

2.17 *Theorem*

Let V be an n-dimensional vector space and W a subspace of V. Then W has finite dimension and any basis of W can be extended to a basis for V by adding in more elements. So if $W \neq V$, then $\dim W < \dim V$.

Proof:
We suppose WLOG that $W \neq \{\mathbf{0}\}$ or V, which is easy. Otherwise, let $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ be an independent set in W chosen to have as many elements as possible. (At most n, since $\dim V = n$.) We claim it's a basis for W.
For any $\mathbf{w} \in W$, the set $\{\mathbf{w}, \mathbf{w}_1, \ldots, \mathbf{w}_k\}$ is not independent, so
$a_1\mathbf{w}_1 + \ldots + a_k\mathbf{w}_k + b\mathbf{w} = \mathbf{0}$, say, with not all the coefficients being zero. But b can't be 0, since $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ is independent. So we can write $\mathbf{w}$ as a combination of $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$, and so they span W, and hence form a basis for it.

Now let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for V. By (2.10) we can replace k of the $\mathbf{v}$'s by $\mathbf{w}$'s and it still spans V. This is the same as extending $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ to a basis of n elements, since any n-element spanning set for V is a basis, by (2.15).

$\square$

2.18 *Example*

Let $W = \{(x,y,z) \in \mathbb{R}^3 : x - 2y + 4z = 0\}$, with general solution $z = a$, $y = b$, $x = 2b - 4a$, i.e., $(2b-4a, b, a) = a(-4,0,1) + b(2,1,0)$.

We can extend it to a basis for $\mathbb{R}^3$ by adding in something chosen from the basis $\{(1,0,0),(0,1,0),(0,0,1)\}$. Indeed, $(1,0,0)$ isn't in the subspace, so that will do.

Thus $\{(-4,0,1),(2,1,0),(1,0,0)\}$ is a basis of $\mathbb{R}^3$ containing the basis for W.

Recall from (1.6) that if $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ are vectors in V, then $\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ consists of all linear combinations $a_1\mathbf{v}_1 + \ldots + a_k\mathbf{v}_k$, and is the smallest subspace containing S.

2.19 *Theorem*

Let V be a finite-dimensional vector space and U a subspace of V. Then there is a subspace W such that $V = U \oplus W$.

We call W a *complement* for U.
Proof: If $U = V$, take $W = \{\mathbf{0}\}$, and if $U = \{\mathbf{0}\}$, take $W = V$.

Otherwise, U has a basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$, which can be extended to a basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{w}_{k+1}, \ldots, \mathbf{w}_n\}$ of V. Let $W = \operatorname{span}\{\mathbf{w}_{k+1}, \ldots, \mathbf{w}_n\}$. We claim that $V = U \oplus W$.

If $\mathbf{v} \in V$, we can write

$$v = \underbrace{a_1\mathbf{u}_1 + \ldots + a_k\mathbf{u}_k}_{\text{in } U} + \underbrace{a_{k+1}\mathbf{w}_{k+1} + \ldots + a_n\mathbf{w}_n}_{\text{in } W}$$

for some $a_1, \ldots, a_n \in F$. So $V = U + W$.

Also $U \cap W = \{\mathbf{0}\}$ (why?) so we have uniqueness, i.e., a direct sum $V = U \oplus W$. $\square$

LECTURE 7

2.20 *Example*

Let $V = \mathbb{R}^3$ and $U = \{a(1,2,3) + b(1,0,6) : a, b \in \mathbb{R}\}$, a plane through $\mathbf{0}$.
For W we can take any line not in the plane that passes through 0, e.g. the x-axis $\{c(1,0,0) : c \in \mathbb{R}\}$. The complement is not unique.

3 Linear mappings

3.1 *Definition*

Let U, V be vector spaces over F. Then a mapping $T : U \to V$ is called a *linear mapping*, or *linear transformation*, if:

(i) $T(\mathbf{u}_1 + \mathbf{u}_2) = T(\mathbf{u}_1) + T(\mathbf{u}_2)$ for all $\mathbf{u}_1, \mathbf{u}_2 \in U$;

(ii) $T(a\mathbf{u}) = aT(\mathbf{u})$ for all $a \in F$ and $\mathbf{u} \in U$.

3.2 *Examples*

(i) Let A be an $m \times n$ matrix of real numbers. Then we define $T : \mathbb{R}^n \to \mathbb{R}^m$ by $\mathbf{y} = T(\mathbf{x}) = A\mathbf{x}$, i.e.,

$$\underset{(m\times 1)}{\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}} = \underset{(m\times n)}{\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}} \underset{(n\times 1)}{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}}.$$

This is linear, and for $m = n = 3$ it includes rotations and reflections.

(ii) The identity mapping on any vector space.

(iii) $D : P_n \to P_{n-1}$, with $Dp = \frac{dp}{dt}$.

(iv) $T : P_n \to F$, with $Tp = \int_0^1 p(t)dt$.

We shall see that for finite-dimensional vector spaces, all linear mappings can be represented by matrices, once we have chosen bases for the spaces U and V involved.

3.3 *Definition*

Let $T : U \to V$ be a linear mapping between vector spaces.
The *null-space*, or *kernel*, of T is $\ker T = \{\mathbf{u} \in U : T(\mathbf{u}) = \mathbf{0}\}$, and is a subset of U.
The *image*, or *range*, of T is $\operatorname{im} T = T(U) = \{T(\mathbf{u}) : \mathbf{u} \in U\}$, and is a subset of V.

Example. Take $T : \mathbb{R}^3 \to \mathbb{R}^4$, defined by $T(x, y, z) = (x, y, x + y, x - y)$ (which is linear). Then

$$\ker T = \{(x, y, z) \in \mathbb{R}^3 : x = y = x + y = x - y = 0\} = \{(0, 0, z) : z \in \mathbb{R}\},$$

and

$$\operatorname{im} T = \{(x, y, x + y, x - y) : x, y \in \mathbb{R}\} = \{x(1, 0, 1, 1) + y(0, 1, 1, -1) : x, y \in \mathbb{R}\}.$$

3.4 *Proposition*

Let $T : U \to V$ be a linear mapping between vector spaces. Then $\ker T$ is a subspace of U and $\operatorname{im} T$ is a subspace of V.

Proof: (i) Start with $\ker T$. Note that $T(\mathbf{0}) = T(\mathbf{0} + \mathbf{0}) = T(\mathbf{0}) + T(\mathbf{0})$, by linearity, and this shows that $T(\mathbf{0}) = \mathbf{0}$. So $\mathbf{0} \in \ker T$.
If $\mathbf{u}_1, \mathbf{u}_2 \in \ker T$, then $T(\mathbf{u}_1 + \mathbf{u}_2) = T(\mathbf{u}_1) + T(\mathbf{u}_2)$ (linearity), which equals $\mathbf{0}+\mathbf{0}$ or $\mathbf{0}$. So $\mathbf{u}_1 + \mathbf{u}_2 \in \ker T$.
If $\mathbf{u} \in \ker T$ and $a \in F$, then $T(a\mathbf{u}) = aT(\mathbf{u})$ (linearity), which equals $a\mathbf{0}$ or $\mathbf{0}$. So $a\mathbf{u} \in \ker T$.
Hence $\ker T$ is a subspace of U.

(ii) Since $T(\mathbf{0}) = \mathbf{0}$, we also have $\mathbf{0} \in \operatorname{im} T$.
If $\mathbf{v}_1, \mathbf{v}_2 \in \operatorname{im} T$, then there exist $\mathbf{u}_1, \mathbf{u}_2 \in U$ such that $\mathbf{v}_1 = T(\mathbf{u}_1)$ and $\mathbf{v}_2 = T(\mathbf{u}_2)$. Then $\mathbf{v}_1 + \mathbf{v}_2 = T(\mathbf{u}_1) + T(\mathbf{u}_2) = T(\mathbf{u}_1 + \mathbf{u}_2)$, so it lies in $\operatorname{im} T$.
Likewise, if $\mathbf{v} \in \operatorname{im} T$, then there exists $\mathbf{u} \in U$ such that $\mathbf{v} = T(\mathbf{u})$, and then $a\mathbf{v} = aT(\mathbf{u}) = T(a\mathbf{u})$, so it lies in $\operatorname{im} T$.
Hence $\operatorname{im} T$ is a subspace of V. □

3.5 *Definition*

For $T : U \to V$ linear, the *nullity* of T is $\dim(\ker T)$, and written $n(T)$.
The *rank* of T is $\dim(\operatorname{im} T)$, and written $r(T)$.

In the example of $T : \mathbb{R}^3 \to \mathbb{R}^4$ we have $n(T) = 1$ and $r(T) = 2$.

LECTURE 8

3.6 *Theorem*

Let U, V be vector spaces over F and $T : U \to V$ linear. If U is finite-dimensional, then $r(T) + n(T) = \dim U$.

Proof: If $U = \{\mathbf{0}\}$, this is clear, so assume $\dim U \geq 1$.
Choose a basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ for $\ker T$ and extend it to a basis
$S = \{\mathbf{w}_1, \ldots, \mathbf{w}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ for U.
(If $k = n$ already, then T is the zero map, and the result is clear.)
We claim that $\{T(\mathbf{u}_{k+1}), \ldots, T(\mathbf{u}_n)\}$ is a basis for $\operatorname{im} T$.

Independence. If $a_{k+1}T(\mathbf{u}_{k+1}) + \ldots + a_n T(\mathbf{u}_n) = \mathbf{0}$, then
$a_{k+1}\mathbf{u}_{k+1} + \ldots + a_n\mathbf{u}_n \in \ker T$.
So $a_{k+1}\mathbf{u}_{k+1} + \ldots + a_n\mathbf{u}_n = a_1\mathbf{w}_1 + \ldots + a_k\mathbf{w}_k$ for some $a_1, \ldots, a_k \in F$. This gives a linear relation between elements of S, and so since S is independent, we conclude that $a_{k+1} = \ldots = a_n = 0$.

Spanning. If $\mathbf{v} \in \operatorname{im} T$, then $\mathbf{v} = T\mathbf{u}$ for some $\mathbf{u} \in U$, and we can find $b_1, \ldots, b_n \in F$ such that $\mathbf{u} = b_1\mathbf{w}_1 + \ldots + b_k\mathbf{w}_k + b_{k+1}\mathbf{u}_{k+1} + \ldots + b_n\mathbf{u}_n$, using our basis for U.
Apply T, and we get $\mathbf{v} = b_{k+1}T(\mathbf{u}_{k+1}) + \ldots + b_nT(\mathbf{u}_n)$, since the $T(\mathbf{w}_i)$ are all 0. So the set spans $\operatorname{im} T$.

Now $r(T) = n - k$, and $n(T) = k$, and indeed $r(T) + n(T) = n = \dim U$.

□

For example, if $T : \mathbb{R}^4 \to \mathbb{R}^2$ is defined by $T(x, y, z, w) = (x + y, 3x + 3y)$, then $\ker T$ is all solutions to $x + y = 3x + 3y = 0$, i.e., parametrised by $(a, -a, b, c)$ and $n(T) = 3$.
Likewise, $\operatorname{im} T$ is parametrised as $(d, 3d)$, and $r(T) = 1$.
Then $r(T) + n(T) = 4 = \dim \mathbb{R}^4$.

4 Linear mappings and matrices

4.1 *Definition*

Let $\mathbf{v}_1, \ldots \mathbf{v}_n$ be a basis of a vector space V and let $\mathbf{v} = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n$. We call $a_1, \ldots, a_n$ the *coordinates* of $\mathbf{v}$ with respect to $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

4.2 *Example*

If $\mathbf{e}_1 = (1, 0, \ldots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \ldots, 0)$, $\ldots$, $\mathbf{e}_n = (0, 0, \ldots, 0, 1)$ is the standard basis of $\mathbb{R}^n$, then $\mathbf{x} = (x_1, \ldots, x_n)$ has coordinates $x_1, \ldots, x_n$.

If $\mathbf{v}_1 = (1, 2)$ and $\mathbf{v}_2 = (-2, 1)$ are given as a basis of $\mathbb{R}^2$, then
$\mathbf{x} = (x, y) = a(1, 2) + b(-2, 1)$ implies that $a - 2b = x$ and $2a + b = y$, so that

$$a = \frac{x + 2y}{5} \qquad \text{and} \qquad b = \frac{-2x + y}{5}$$

are the coordinates of $\mathbf{x}$ with respect to $\mathbf{v}_1, \mathbf{v}_2$.

4.3 *Definition (based on (3.2))*

Let U and V be finite-dimensional vector spaces over a field F, with bases $\mathbf{u}_1, \ldots, \mathbf{u}_n$ and $\mathbf{v}_1, \ldots, \mathbf{v}_m$, and let $T : U \to V$ be a linear mapping.
T is represented by an $m \times n$ matrix A with respect to the given basis if, whenever $\mathbf{x} \in U$ has coordinates $x_1, \ldots, x_n$, then $T(\mathbf{x}) \in V$ has coordinates $y_1, \ldots, y_m$, where

$$\underset{m \times 1}{\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}} = \underset{m \times n}{A} \; \underset{n \times 1.}{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}}$$

Note $y_i = \sum_{j=1}^n a_{ij}x_j$ for $i = 1, \ldots, m$, where A has entries $(a_{ij})_{i=1}^m{}_{j=1}^n$.

So clearly $T : \mathbb{R}^n \to \mathbb{R}^m$ given by left multiplication by A is represented by A if we use the standard basis.

4.4 *Proposition*

Let $U, V, \mathbf{u}_1, \ldots, \mathbf{u}_n, \mathbf{v}_1, \ldots, \mathbf{v}_m$ be as in (4.3). Every map T such that the coordinates of $T(\mathbf{x})$ with respect to the $\mathbf{v}$'s is given by

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

(where $\mathbf{x}$ has coordinates $x_1, \ldots, x_n$ with respect to the $\mathbf{u}$'s) is linear.

Proof: Suppose that $\mathbf{x}$ has coordinates $x_1, \ldots, x_n$ and $\mathbf{x}'$ has coordinates $x'_1, \ldots, x'_n$. Then $T(\mathbf{x} + \mathbf{x}')$ has coordinates

$$A \begin{pmatrix} x_1 + x'_1 \\ \vdots \\ x_n + x'_n \end{pmatrix} = A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + A \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}$$

which are the coordinates of $T(\mathbf{x})$ added to the coordinates of $T(\mathbf{x}')$.
So $T(\mathbf{x} + \mathbf{x}') = T(\mathbf{x}) + T(\mathbf{x}')$.

Similarly, $T(a\mathbf{x}) = aT(\mathbf{x})$ by looking at coordinates.

□

How do we find the matrix of $T : U \to V$ if we are given bases $\mathbf{u}_1, \ldots, \mathbf{u}_n$ and $\mathbf{v}_1, \ldots, \mathbf{v}_m$?

Note that $T(\mathbf{u}_1) \in V$, so can be written as a combination of the $\mathbf{v}$'s. Indeed, to find that combination, if there is a matrix A representing T, we must have

$$A \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

if $T(\mathbf{u}_1) = b_1\mathbf{v}_1 + \ldots + b_m\mathbf{v}_m$. That is, $\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$ is the first column of A. Similarly,

$$A \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix}$$

if $T(\mathbf{u}_2) = c_1\mathbf{v}_1 + \ldots + c_m\mathbf{v}_m$; and so on.

LECTURE 9

Example. Suppose that $T : U \to V$, where U has basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ and V has basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, and that
$T(\mathbf{u}_1) = 3\mathbf{v}_1 + 4\mathbf{v}_2 + 5\mathbf{v}_3$, while $T(\mathbf{u}_2) = \mathbf{v}_1 + \mathbf{v}_2 + 9\mathbf{v}_3$. Then we fill in the columns to get $A = \begin{pmatrix} 3 & 1 \\ 4 & 1 \\ 5 & 9 \end{pmatrix}$, and note that $A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}$, while $A \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 9 \end{pmatrix}$.

Note that the identity mapping $I : U \to U$ with $I(\mathbf{u}) = \mathbf{u}$ corresponds to the identity matrix $\begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ldots & \vdots \\ \vdots & \vdots & \ddots & \ldots & 0 \\ 0 & 0 & \ldots & \ldots & 1 \end{pmatrix}$ of size $\dim U$, written I or I_n if it is $n \times n$.

This gives us:

4.5 *Proposition*

The matrix A representing T with respect to $\mathbf{u}_1, \ldots, \mathbf{u}_n$ and $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is the one whose ith column is $\begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{pmatrix}$, where
$T(\mathbf{u}_i) = a_{1i}\mathbf{v}_1 + a_{2i}\mathbf{v}_2 + \ldots + a_{mi}\mathbf{v}_m = \sum_{j=1}^{m} a_{ji}\mathbf{v}_j$.

Proof: For a typical vector $\mathbf{x} = x_1\mathbf{u}_1 + \ldots + x_n\mathbf{u}_n$, we have

$$\begin{aligned} T(\mathbf{x}) &= \sum_{i=1}^{n} x_i T(\mathbf{u}_i) \qquad \text{(by linearity)} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i a_{ji}\mathbf{v}_j = \sum_{j=1}^{m} y_j v_j, \end{aligned}$$

where

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

for $j = 1, \ldots, m$, i.e., $y_j = \sum_{i=1}^{n} a_{ji}x_i$.

□

4.6 *Example*

Find the matrix of the linear mapping $T : \mathbb{R}^3 \to \mathbb{R}^2$, with
$T(x, y, z) = (x + y + z, 2x + 2y + 2z)$,
(i) with respect to the standard bases of $\mathbb{R}^3$ (call it A);
(ii) with respect to the bases $\{(1, 0, 0), (1, -1, 0), (0, 1, -1)\}$ of $\mathbb{R}^3$ and $\{(1, 2), (1, 0)\}$ of $\mathbb{R}^2$ (call it B).

Solution. (i) $T(1, 0, 0) = (1, 2)$, $T(0, 1, 0) = (1, 2)$ and $T(0, 0, 1) = (1, 2)$, so fill in columns to get $A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix}$.
(ii)

$$\begin{aligned} T(1, 0, 0) &= (1, 2) = 1(1, 2) + 0(1, 0) \\ T(1, -1, 0) &= (0, 0) = 0(1, 2) + 0(1, 0) \\ T(0, 1, -1) &= (0, 0) = 0(1, 2) + 0(1, 0) \end{aligned}$$

Filling in columns we get $B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

4.7 *Theorem*

Let $T : U \to V$ be a linear mapping between vector spaces, and suppose that $\dim U = n$, $\dim V = m$. Then we can find bases $\mathbf{u}_1, \ldots, \mathbf{u}_n$ of U and $\mathbf{v}_1, \ldots, \mathbf{v}_m$ of V so that the matrix of T has the *canonical form*

$$A = \begin{pmatrix} I_r & O \\ O & O \end{pmatrix},$$

where $r = \text{rank}(T)$, I_r is the identity matrix of size $r \times r$, and the rest is 0.

Proof: Recall that $\ker T$ has dimension $n - r$, the nullity of T, by (3.6). So take a basis $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n$ of $\ker(T)$ and extend to a basis $\mathbf{u}_1, \ldots, \mathbf{u}_n$ of U.
Let $\mathbf{v}_1 = T(\mathbf{u}_1), \ldots, \mathbf{v}_r = T(\mathbf{u}_r)$.
As in (3.6), $\mathbf{v}_1, \ldots, \mathbf{v}_r$ is a basis of $\operatorname{im} T$ and we can extend to a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of V.

Now $T(\mathbf{u}_1) = \mathbf{v}_1$, so the first column of the matrix will be $\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$; and so on, until $T(\mathbf{u}_r) = \mathbf{v}_r$, so the rth column will be $\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, with the 1 in the rth row. Finally, $T(\mathbf{u}_{r+1}) = \ldots = T(\mathbf{u}_n) = 0$, as these vectors are in the kernel; so the remaining columns are all 0.

$\square$

4.8 *Another example*

Take $T : \mathbb{R}^2 \to \mathbb{R}^2$ with $T(x, y) = (x - 2y, 2x - 4y)$. Now $T(1, 0) = (1, 2)$ and $T(0, 1) = (-2, -4)$, so the matrix is $\begin{pmatrix} 1 & -2 \\ 2 & -4 \end{pmatrix}$ with respect to the standard bases.

LECTURE 10

Now $\ker T$ is all multiples of $(2, 1)$ so take $\mathbf{u}_2 = (2, 1)$ and we can take $\mathbf{u}_1 = (1, 0)$ so that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is a basis for $\mathbb{R}^2$.

Then $T(\mathbf{u}_1) = (1, 2)$, which gives a basis for $\operatorname{im} T$, and so we let $\mathbf{v}_1 = (1, 2)$. Extend with $\mathbf{v}_2 = (0, 1)$ (say) to another basis for $\mathbb{R}^2$.

Now T has the matrix $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ with respect to the bases $\{\mathbf{u}_1, \mathbf{u}_2\}$ and $\{\mathbf{v}_1, \mathbf{v}_2\}$.
Also $r(T) = 1$ and $n(T) = 1$.

4.9 *Proposition*

Let A be an $m \times n$ matrix with real entries. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be defined by $T(\mathbf{x}) = A\mathbf{x}$. Then
$n(T)$ is the dimension of the solution space of the equations $A\mathbf{x} = \mathbf{0}$, and

$r(T)$ is the dimension of the subspace of $\mathbb{R}^m$ spanned by the columns of A.

Proof:
The result on $n(T)$ is just by definition of the kernel.

For $r(T)$, since

$$A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum a_{1j}x_j \\ \sum a_{2j}x_j \\ \vdots \\ \sum a_{mj}x_j \end{pmatrix} = x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + \ldots + x_n \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix},$$

we see that $\operatorname{im} T$ is the span of the columns of A.

□

4.10 *Corollary*

The row rank of a matrix (number of independent rows) equals the column rank (number of independent columns).

Proof: Define $T : \mathbb{R}^n \to \mathbb{R}^m$ by $T(\mathbf{x}) = A\mathbf{x}$. By (4.9), the column rank of A is $r(T)$, which is $n - n(T)$. This is n-[dimension of solution space to $A\mathbf{x} = \mathbf{0}$], i.e., n-[number of free parameters in the solution], which is the number of non-zero rows in the reduced form of A, which is the row rank of A.

□

For example, $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{pmatrix}$ has row rank 1 and column rank 1. The solutions to $A\mathbf{x} = \mathbf{0}$ form a two-dimensional space.

Composition of mappings.

4.11 *Proposition*

Let U, V, W be vector spaces over F and let $T : U \to V$ and $S : V \to W$ be linear mappings. Then ST is a linear mapping from U to W.

Proof: Clearly ST (i.e., T followed by S) maps U into W.
Also $ST(\mathbf{u}_1 + \mathbf{u}_2) = S(T(\mathbf{u}_1) + T(\mathbf{u}_2))$, by linearity of T, and this is $S(T(\mathbf{u}_1)) + S(T(\mathbf{u}_2))$, by linearity of S.
Similarly, we see that $ST(a\mathbf{u}) = S(aT\mathbf{u}) = a(S(T(\mathbf{u}))$, and so ST is linear.

□

4.12 *Example*

Let $U = \mathbb{R}^2$, $V = \mathbb{R}^3$ and $W = \mathbb{R}^4$. We define $T(x_1, x_2) = (x_1 + x_2, x_1, x_2)$ and $S(y_1, y_2, y_3) = (y_1 + y_2, y_1, y_2, y_3)$.
Then $ST(x_1 + x_2) = (2x_1 + x_2, x_1 + x_2, x_1, x_2)$.

4.13 *Proposition*

Let U, V, W be finite-dimensional vector spaces over F with bases $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$, $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_\ell\}$, and let $T : U \to V$ and $S : V \to W$ be linear mappings. Let T be represented by B and S by A with respect to the given bases. Then ST is represented by the matrix AB, i.e.,

$$\begin{array}{ccccc} & T & & S & \\ U & \to & V & \to & W \\ & B & & A & \end{array}$$

Proof: $T(\mathbf{u}_j) = \sum_{i=1}^{m} b_{ij}\mathbf{v}_i$ and $S(\mathbf{v}_i) = \sum_{k=1}^{\ell} a_{ki}\mathbf{w}_k$, as in Proposition 4.5. So

$$\begin{aligned} ST(\mathbf{u}_j) &= \sum_{i=1}^{m} b_{ij} S(\mathbf{v}_i) = \sum_{i=1}^{m}\sum_{k=1}^{\ell} b_{ij} a_{ki} \mathbf{w}_k \\ &= \sum_{k=1}^{\ell} \left(\sum_{i=1}^{m} a_{ki} b_{ij} \right) \mathbf{w}_k \\ &= \sum_{k=1}^{\ell} (AB)_{kj} \mathbf{w}_k. \end{aligned}$$

□

This "explains" the rule for multiplying matrices.

4.14 *Example 4.12 revisited*

T has matrix $B = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$ and S has matrix $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Then $AB = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$, the matrix of ST.

LECTURE 11

Isomorphisms

4.15 *Definition*

An *isomorphism* is a linear mapping $T : U \to V$ of vector spaces over the same field, for which there is an inverse mapping $T^{-1} : V \to U$ satisfying $T^{-1}T = I_U$ and $TT^{-1} = I_V$, where I_U and I_V are the identity mappings on U and V respectively.

4.16 *Theorem*

Two finite-dimensional vector spaces U and V are isomorphic if and only if $\dim U = \dim V$.

Proof: If $T : U \to V$ is an isomorphism, then every $\mathbf{v} \in V$ is in $\operatorname{im} T$, since $\mathbf{v} = T(T^{-1}(\mathbf{v}))$, so we have $\operatorname{im} T = V$ and $\dim U = r(T) + n(T) \geq \dim V$.

Similarly (look at $T^{-1} : V \to U$, which is also an isomorphism), we have $\dim V \geq \dim U$. So $\dim U = \dim V$, and hence $r(T) = \dim U = \dim V$.

Conversely, if $\dim U = \dim V$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ are bases of U and V, then we can define an isomorphism by

$$T(a_1\mathbf{u}_1 + \ldots + a_n\mathbf{u}_n) = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n,$$

for all $a_1, \ldots, a_n \in F$, and clearly $T^{-1}(a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n) = a_1\mathbf{u}_1 + \ldots + a_n\mathbf{u}_n$. □

Example.

Let $U = \mathbb{R}^2$, and let V be the space of all real polynomials p of degree ≤ 2 such that $p(1) = 0$. Clearly $\dim U = 2$. For V, we note that
$V = \{a_0 + a_1t + a_2t^2 : a_0 + a_1 + a_2 = 0\}$. Setting $a_0 = c$ and $a_1 = d$ we have $a_2 = -c - d$, so
$V = \{c + dt + (-c - d)t^2 : c, d \in \mathbb{R}\} = \{c(1 - t^2) + d(t - t^2) : c, d \in \mathbb{R}\}$, with basis $\{1 - t^2, t - t^2\}$.
Hence $\dim V = 2$, and there is an isomorphism between U and V defined by $T(a, b) = T(a(1, 0) + b(0, 1)) = a(1 - t^2) + b(t - t^2)$.

4.17 *Remark*

If A represents T with respect to some given bases of U and V, then A^{-1} represents T^{-1}, since if B is the matrix of T^{-1}, we have:
$BA =$ matrix of $T^{-1}T = I_n$, and
$AB =$ matrix of $TT^{-1} = I_n$, by (4.13).

5 Matrices and change of bases

The idea of this section is to choose bases for U and V so that $T : U \to V$ has the simplest possible matrix, namely $\begin{pmatrix} I_r & O \\ O & O \end{pmatrix}$, as in (4.7). How is this related to the original matrix?

5.1 *Proposition*

Let V be an n-dimensional vector space over F. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of V and $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be any set of n vectors, not necessarily distinct, in V. Then
(i) There is a unique linear mapping $S : V \to V$ such that $S\mathbf{v}_j = \mathbf{w}_j$ for each j.
(ii) There is a unique square matrix P representing S in the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, such that $\mathbf{w}_j = \sum_{i=1}^n p_{ij}\mathbf{v}_i$, for $j = 1, \ldots, n$.
(iii) $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ is a basis of V if and only if P is non-singular, i.e., invertible.

Proof: (i) Define $S(\sum_{j=1}^n x_j\mathbf{v}_j) = \sum_{j=1}^n x_j\mathbf{w}_j$. This is clearly linear and the only possibility.

(ii) We write $\mathbf{w}_j = \sum_{i=1}^n p_{ij}\mathbf{v}_i$ using the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. This determines the matrix P, which is the matrix of S, as in (4.5).

(iii) If $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ is a basis of V, then there's a linear mapping $T : V \to V$ such that $T(\mathbf{w}_j) = \mathbf{v}_j$ for each j. Now $ST = TS = I$ (identity), so that the matrix R of T satisfies $PR = RP = I_n$, i.e., $R = P^{-1}$.

Conversely, if P is non-singular, then P^{-1} represents a linear mapping T such that $T(\mathbf{w}_j) = \mathbf{v}_j$ for each j. But if $\sum_{j=1}^n a_j\mathbf{w}_j = \mathbf{0}$, then, applying T, we get $\sum_{j=1}^n a_j\mathbf{v}_j = \mathbf{0}$, so $a_1 = \ldots = a_n = 0$, as $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis. Hence $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ is independent, and since $\dim V = n$ this means it's a basis, by (2.15).

$\square$

LECTURE 12

5.2 *Theorem*

Let U and V be finite-dimensional vector spaces over F and $T : U \to V$ a linear mapping represented by a matrix A with respect to basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ of U and $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ of V. Then the matrix B representing T with respect to new bases $\{\mathbf{u}'_1, \ldots, \mathbf{u}'_n\}$ of U and $\{\mathbf{v}'_1, \ldots, \mathbf{v}'_m\}$ of V is given by $B = Q^{-1}AP$, where
P is the matrix of the identity mapping on U with respect to the bases $\{\mathbf{u}'_1, \ldots, \mathbf{u}'_n\}$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$, i.e., $\mathbf{u}'_j = \sum_{i=1}^n p_{ij}\mathbf{u}_i$ (so it writes the new basis in terms of the old one), and similarly

Q is the matrix of the identity mapping on V with respect to the bases $\{\mathbf{v}'_1, \ldots, \mathbf{v}'_m\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$, i.e., $\mathbf{v}'_k = \sum_{i=1}^{n} q_{\ell k} \mathbf{v}_\ell$.

Proof: It's a composition of mappings, and hence a product of matrices:

$$\left.\begin{array}{cccc}
\text{Space} & \text{Basis} & \text{Mapping} & \text{Matrix} \\
U & \{\mathbf{u}'_1, \ldots, \mathbf{u}'_n\} & & \\
 & & I\downarrow & P \\
U & \{\mathbf{u}_1, \ldots, \mathbf{u}_n\} & & \\
 & & T\downarrow & A \\
V & \{\mathbf{v}_1, \ldots, \mathbf{v}_m\} & & \\
 & & I\downarrow & Q^{-1} \\
V & \{\mathbf{v}'_1, \ldots, \mathbf{v}'_m\} & &
\end{array}\right\} B,$$

So $B = Q^{-1}AP$.

□

5.3 *Definition*

Two $m \times n$ matrices A, B with entries in F are *equivalent*, if there are non-singular square matrices P, Q with entries in F such that the product $Q^{-1}AP$ is defined and equals B. (So P must be $n \times n$ and Q is $m \times m$.)
Writing $R = Q^{-1}$, we could also say $B = RAP$, with R, P nonsingular.
If A and B are equivalent, we write $A \equiv B$.

5.4 *Proposition*

Equivalence of matrices is an *equivalence relation*; i.e., for $m \times n$ matrices over F, we have

$$A \equiv A, \qquad A \equiv B \implies B \equiv A, \qquad [A \equiv B, \quad B \equiv C] \implies A \equiv C.$$

Proof: (i) $A = I_m A I_n$, so $A \equiv A$.

(ii) If $A \equiv B$, so $B = Q^{-1}AP$, then $A = QBP^{-1} = (Q^{-1})^{-1}BP^{-1}$, so $B \equiv A$.

(iii) If $A \equiv B$ and $B \equiv C$, say $B = Q^{-1}AP$ and $C = S^{-1}BR$, then
$C = S^{-1}Q^{-1}APR = (QS)^{-1}A(PR)$, so $A \equiv C$.

□

5.5 *Theorem*

Let U, V be vector spaces over F, with $\dim U = n$ and $\dim V = m$, and let A be an $m \times n$ matrix. Then

(i) Given bases $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ of U and V, there is a linear mapping T that is represented by A with respect to these bases.

(ii) An $m \times n$ matrix B satisfies $A \equiv B$ if and only if B represents T with respect to *some* bases of U and V.

(iii) There is a unique matrix C of the form $\begin{pmatrix} I_r & O \\ O & O \end{pmatrix}$, such that $A \equiv C$. Moreover $r = \text{rank}(T)$.

Proof: (i) For $\mathbf{u} = x_1\mathbf{u}_1 + \ldots + x_n\mathbf{u}_n$ we write $T(\mathbf{u}) = y_1\mathbf{v}_1 + \ldots + y_m\mathbf{v}_m$, where $\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

(ii) This follows from (5.2).

(iii) This follows from (4.7).

□

5.6 *Example*

Take $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 2 \end{pmatrix}$. Find r and nonsingular P and Q such that $Q^{-1}AP = \begin{pmatrix} I_r & O \\ O & O \end{pmatrix}$. Note that P must be 3×3 and Q must be 2×2.

Solution. Row and column reduce to get it into canonical form. We'll do rows first.

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where we did $r_2 - r_1$ and then $r_2/2$. Next, columns.

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

doing $c_2 - c_1$ and $c_2 \leftrightarrow c_3$.

For Q^{-1} start with the 2×2 identity matrix I_2, and do the same row operations.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

For P start with I_3 and do the same column operations.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

We can get Q by inverting Q^{-1}, and the answer is $\begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}$.

LECTURE 13

Alternatively. From first principles, we have $T(x, y, z) = (x + y, x + y + 2z)$, so the kernel consists of all vectors with $z = 0$ and $x = -y$, i.e., has a basis $(-1, 1, 0)$.
Extend to a basis for $\mathbb{R}^3$, say by adding in $(1, 0, 0)$ and $(0, 0, 1)$.
Fill in the basis vectors as columns. So $P = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$, as before.
Also $T(1, 0, 0) = (1, 1)$ and $T(0, 0, 1) = (0, 2)$, which already gives a basis for $\mathbb{R}^2$.
Fill in columns so $Q = \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}$, also as before.

6 Linear mappings from a vector space to itself

Now for $T : V \to V$, we shall see what can be done using only one basis, say $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. A matrix A represents T if and only if the coordinates of $T(\mathbf{v})$ with respect to the basis are $\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, where $\mathbf{v}$ has coordinates $x_1, \ldots, x_n$; i.e.,

$$[\mathbf{v} = x_1\mathbf{v}_1 + \ldots + x_n\mathbf{v}_n] \qquad \Longrightarrow \qquad [T(\mathbf{v}) = y_1\mathbf{v}_1 + \ldots + y_n\mathbf{v}_n].$$

Here A is $n \times n$ (square).

Also $T(\mathbf{v}_j) = a_{1j}\mathbf{v}_1 + \ldots + a_{nj}\mathbf{v}_n$, and $\begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix}$ is the jth column of A.

Consider the linear mapping $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $\mathbf{x} \mapsto A\mathbf{x}$ with $A = \begin{pmatrix} 1 & 1 \\ -2 & 4 \end{pmatrix}$.
If we use the basis consisting of $\mathbf{v}_1 = (1, 1)$ and $\mathbf{v}_2 = (1, 2)$, then
$T(\mathbf{v}_1) = (2, 2) = 2\mathbf{v}_1$ and $T(\mathbf{v}_2) = (3, 6) = 3\mathbf{v}_2$.
Hence with respect to this basis T has the diagonal matrix $B = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$. Can we always represent a linear mapping by such a simple matrix?

6.1 *Theorem*

Let V be an n-dimensional vector space over F, and $T : V \to V$ a linear mapping represented by the matrix A with respect to the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. Then T is represented by the matrix B with respect to the basis $\{\mathbf{v}'_1, \ldots, \mathbf{v}'_n\}$ if and only if $B = P^{-1}AP$, where $P = (p_{ij})$ is nonsingular and $\mathbf{v}'_j = \sum_{i=1}^{n} p_{ij}\mathbf{v}_i$, i.e., P is the matrix of the identity map on B with respect to the bases $\{\mathbf{v}'_1, \ldots, \mathbf{v}'_n\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$.

Proof: This is just Theorem 5.2 with $Q = P$.

□

In our example, $P = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$, filling in $\mathbf{v}_1$ and $\mathbf{v}_2$ as the columns.

6.2 *Definition*

Two $n \times n$ matrices A and B over F are *similar* or *conjugate* if there is a non-singular square matrix P with $B = P^{-1}AP$. We write $A \sim B$. This happens if they represent the same linear mapping with respect to two bases.

Thus, in our example, $\begin{pmatrix} 1 & 1 \\ -2 & 4 \end{pmatrix} \sim \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$.

6.3 *Proposition*

Similarity is an equivalence relation on the set of $n \times n$ matrices over F, i.e.,

$$A \sim A, \qquad A \sim B \implies B \sim A, \qquad [A \sim B, \quad B \sim C] \implies A \sim C.$$

Proof: (i) $A = I^{-1}AI$, so $A \sim A$.
(ii) If $A \sim B$, i.e., $B = P^{-1}AP$, then $A = PBP^{-1} = M^{-1}BM$, with $M = P^{-1}$. So $B \sim A$.
(iii) If $A \sim B$ and $B \sim C$, so that $B = P^{-1}AP$ and $C = Q^{-1}BQ$, then $C = Q^{-1}P^{-1}APQ = N^{-1}AN$, with $N = PQ$. So $A \sim C$.

□

6.4 *Example*

Take $T : \mathbb{R}^2 \to \mathbb{R}^2$ with $T(\mathbf{v}) = A\mathbf{v}$, where $A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = 2I_2$. Now $P^{-1}AP = P^{-1}(2I_2)P = P^{-1}(2P) = 2P^{-1}P = 2I_2$, for every P, so that whatever basis we use, T has to have the matrix A. Note $T(\mathbf{x}) = 2\mathbf{x}$ for all $\mathbf{v}$, so that 2 is an "eigenvalue" – this turns out to be the key.

6.5 *Definition*

Let V be a vector space over F and $T : V \to V$ a linear mapping. An *eigenvector* of T is an element $\mathbf{v} \neq \mathbf{0}$ such that $T(\mathbf{v}) = \lambda\mathbf{v}$ for some $\lambda \in F$. The scalar λ is then called an *eigenvalue*.

For a matrix A with entries in F, then $\lambda \in F$ is an eigenvalue, if there is a non-zero $\mathbf{x} \in F^n$ such that $A\mathbf{x} = \lambda\mathbf{x}$, and then $\mathbf{x}$ is an eigenvector of A.
For the mapping $T : F^n \to F^n$ given by $T(\mathbf{x}) = A\mathbf{x}$, this is the same definition of course.

In our example, $\lambda = 2$ and 3 were eigenvalues, with eigenvectors $(1, 1)$ and $(1, 2)$ respectively.

LECTURE 14

6.6 *Proposition*

A scalar $\lambda \in F$ is an eigenvalue of an $n \times n$ matrix A if and only if λ satisfies the *characteristic equation*,

$$\chi(\lambda) = \det(\lambda I_n - A) = 0,$$

which is a polynomial of degree n.

N.B. Some people use $\det(A - \lambda I_n)$ as the definition of $\chi(\lambda)$. We accept either definition, as they only differ by a factor of $(-1)^n$, and so have the same roots.

Proof: A matrix M is invertible if and only if $\det M \neq 0$.
For $(\det M)(\det M^{-1}) = \det(MM^{-1}) = \det I_n = 1$, so an invertible matrix has non-zero determinant. Conversely, if $\det M \neq 0$, then M can't be reduced to a matrix with a row of zeroes, so it has rank n, so is invertible.

Now, thinking of A as giving a linear mapping on F^n, as usual, we have:
λ is an eigenvalue of A
$\Leftrightarrow A - \lambda I$ has a non-zero kernel
$\Leftrightarrow A - \lambda I$ isn't invertible
$\Leftrightarrow \det(A - \lambda I) = 0$.

□

In our example, $A = \begin{pmatrix} 1 & 1 \\ -2 & 4 \end{pmatrix}$, and

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} 1-\lambda & 1 \\ -2 & 4-\lambda \end{vmatrix} \\ &= (1-\lambda)(4-\lambda) - (-2)(1) \\ &= \lambda^2 - 5\lambda + 4 + 2 = \lambda^2 - 5\lambda + 6 = (\lambda - 2)(\lambda - 3), \end{aligned}$$

and the eigenvalues are 2 and 3.

6.7 *Proposition*

Similar matrices have the same characteristic equation, and hence the same eigenvalues.

Proof: Let A be an $n \times n$ matrix and $A \sim B$, so $B = P^{-1}AP$, where P is nonsingular.

Then $\det(\lambda I_n - B) = \det(\lambda P^{-1} I_n P - P^{-1}AP) = \det P^{-1}(\lambda I_n - A)P$
$= (\det P^{-1}) \det(\lambda I_n - A)(\det P) = \det(\lambda I_n - A)$, since $\det P^{-1} = (\det P)^{-1}$.

□

6.8 *Proposition*

Let V be an n-dimensional vector space over a field F, and $T : V \to V$ a linear mapping. Let A represent T with respect to the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of V. Then A and T have the same eigenvalues.

Proof: Well, $\lambda \in F$ is an eigenvalue of A

$\Leftrightarrow A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ for some $x_1, \ldots, x_n$ not all 0 in F

$\Leftrightarrow T(x_1\mathbf{v}_1 + \ldots + x_n\mathbf{v}_n) = \lambda(x_1\mathbf{v}_1 + \ldots + x_n\mathbf{v}_n)$ for some $x_1, \ldots, x_n$ not all 0 in F

$\Leftrightarrow T(\mathbf{v}) = \lambda\mathbf{v}$ for some $\mathbf{v} \neq \mathbf{0}$ in V

$\Leftrightarrow \lambda$ is an eigenvalue of T.

□

If we can find a basis of eigenvectors, the matrix of T has a particularly nice form. We'll prove one little result, then see some examples.

6.9 *Proposition*

Let V be an n-dimensional vector space over a field F, and let $T : V \to V$ be a linear mapping. Suppose that $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of eigenvectors of T. Then, with respect to this basis, T is represented by a diagonal matrix whose diagonal entries are eigenvalues of T.

Proof: We have $T(\mathbf{v}_j) = \lambda_j\mathbf{v}_j$, where λ_j is the appropriate eigenvalue (for $j = 1, 2, \ldots, n$).
Recall that $T(\mathbf{v}_j) = \sum_{i=1}^{n} a_{ij}\mathbf{v}_i$ if A is the matrix representing T (these numbers are the jth column of A).

Now $a_{ij} = \begin{cases} \lambda_j & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases}$ so $A = \begin{pmatrix} \lambda_1 & \ldots & 0 \\ 0 & \ddots & 0 \\ 0 & \ldots & \lambda_n \end{pmatrix}$.

□

6.10 *Example*

Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\mathbf{x}) = \begin{pmatrix} 11 & -2 \\ 3 & 4 \end{pmatrix} \mathbf{x}$, and take $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.
Now
$T(\mathbf{v}_1) = \begin{pmatrix} 11 & -2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 5 \\ 15 \end{pmatrix} = 5 \begin{pmatrix} 1 \\ 3 \end{pmatrix}$, and
$T(\mathbf{v}_2) = \begin{pmatrix} 11 & -2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 20 \\ 10 \end{pmatrix} = 10 \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.
So using $\{\mathbf{v}_1, \mathbf{v}_2\}$ the matrix of T is $\begin{pmatrix} 5 & 0 \\ 0 & 10 \end{pmatrix}$.

However, not all matrices have a basis of eigenvectors.

6.11 *Example*

Take $T : \mathbb{R}^2 \to \mathbb{R}^2$, defined by $T(\mathbf{x}) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \mathbf{x}$. This is a rotation of the plane through 90° anti-clockwise.
Now $\det \left\{ \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\} = \begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 + 1$, so no real eigenvalues (roots of $\lambda^2+1 = 0$), so no eigenvectors in $\mathbb{R}^2$ (which is also obvious geometrically).

FACT. Over $\mathbb{C}$ every polynomial can be factorized into linear factors, and so has a full set of complex roots. This is the Fundamental Theorem of Algebra (see MATH 2090). So we can always find an eigenvalue.

LECTURE 15

6.12 *Example*

Consider $T : \mathbb{C}^2 \to \mathbb{C}^2$ (or indeed $\mathbb{R}^2 \to \mathbb{R}^2$) defined by $T(\mathbf{x}) = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \mathbf{x}$.

$$\det \left(\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) = \begin{vmatrix} 2-\lambda & 1 \\ 0 & 2-\lambda \end{vmatrix} = (\lambda - 2)^2,$$

so 2 is the only eigenvalue. Solving $T(\mathbf{x}) = 2\mathbf{x}$ gives

$$\begin{aligned} \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= 2 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \qquad \text{i.e.,} \\ \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \end{aligned}$$

so $x_2 = 0$, x_1 arbitrary, and the eigenvectors are $\{\begin{pmatrix} a \\ 0 \end{pmatrix} : a \in F, a \neq 0\}$, with $F = \mathbb{R}$ or $\mathbb{C}$.

So there is no basis of eigenvectors, and $\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$ is not similar to a diagonal matrix.

6.13 *Theorem*

Let V be a vector space over F, let $T : V \to V$ be linear, and suppose that $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ are eigenvectors of T corresponding to **distinct** eigenvalues $\{\lambda_1, \ldots, \lambda_k\}$. Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a linearly independent set. If $k = \dim V$, then it's a basis.

Proof: Suppose that $\sum_{i=1}^k a_i \mathbf{v}_i = \mathbf{0}$, and apply $(T - \lambda_2 I)(T - \lambda_3 I) \ldots (T - \lambda_k I)$ to both sides.
Each term is sent to zero except for the first which becomes $(\lambda_1 - \lambda_2) \ldots (\lambda_1 - \lambda_k) a_1 \mathbf{v}_1 = \mathbf{0}$.
But $\mathbf{v}_1 \neq \mathbf{0}$, and the other factors are nonzero, so $a_1 = 0$.
Hence $a_2\mathbf{v}_2 + \ldots + a_k\mathbf{v}_k = \mathbf{0}$. Applying $(T - \lambda_3 I) \ldots (T - \lambda_k I)$, we deduce similarly that $a_2 = 0$. Continuing, we see that the a_j are all 0 and the set is independent. Finally, if we have k independent vectors in a k-dimensional space, then it is automatically a basis, by (2.15).

□

6.14 *Corollary*

(i) Let V be an n-dimensional vector space over a field F, and $T : V \to V$ a linear transformation. If T has n distinct eigenvalues in F, then V has a basis of eigenvectors of T, and T can be represented by a diagonal matrix, whose diagonal entries are the eigenvalues; this is unique up to reordering the eigenvalues.

(ii) Let A be an $n \times n$ matrix with entries in $F = \mathbb{R}$ or $\mathbb{C}$. If the characteristic polynomial of A has n distinct roots in F, then A is similar to a diagonal matrix over F.

Proof: (i) Take $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ eigenvectors of T corresponding to different eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$, say. By (6.13), they are independent, so since there are n of them, they are a basis for the n-dimensional space V (see Theorem 2.15). By (6.9), T is represented by $\begin{pmatrix} \lambda_1 & \ldots & 0 \\ 0 & \ddots & 0 \\ 0 & \ldots & \lambda_n \end{pmatrix}$ with respect to this basis.

(ii) Immediate – let $V = F^n$ and $T(\mathbf{x}) = A\mathbf{x}$.

□

A matrix similar to a diagonal matrix is called *diagonalisable*. Not all matrices are, e.g., if $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, then $A^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, so if $P^{-1}AP = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}$, then $P^{-1}A^2P = (P^{-1}AP)(P^{-1}AP) = \begin{pmatrix} a_1^2 & 0 \\ 0 & a_2^2 \end{pmatrix}$. Since $A^2 = O$, we have $a_1 = a_2 = 0$. This means that $A = P \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} P^{-1} = O$, which is a contradiction.

Of course $I_n = \begin{pmatrix} 1 & \dots & 0 \\ 0 & \ddots & 0 \\ 0 & \dots & 1 \end{pmatrix}$ is diagonalisable, even though it has repeated eigenvalues, 1, so (6.14) isn't the only way a matrix can be diagonalisable.

When $T(\mathbf{x}) = A\mathbf{x}$ and A is diagonalisable, then we calculate P such that $D = P^{-1}AP$ is diagonal by expressing the eigenvectors of A as its columns.

6.15 *Example*

Let $A = \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix}$. Find D such that $A \sim D$, and P such that $D = P^{-1}AP$.

Eigenvalues.

$$\begin{aligned} \det(\lambda I - A) = \begin{vmatrix} \lambda - 1 & -3 \\ -4 & \lambda - 2 \end{vmatrix} &= (\lambda - 1)(\lambda - 2) - 12 \\ &= \lambda^2 - 3\lambda - 10 = (\lambda - 5)(\lambda + 2), \end{aligned}$$

so the eigenvalues are 5 and -2, and we can take $D = \begin{pmatrix} 5 & 0 \\ 0 & -2 \end{pmatrix}$.

For $\lambda = 5$ we solve $(A - 5I)\mathbf{x} = \mathbf{0}$, or $\begin{pmatrix} -4 & 3 \\ 4 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, i.e., $(x_1, x_2) = (\frac{3}{4}a, a)$ for $a \in \mathbb{R}$. So we take $\mathbf{v}_1 = (3, 4)$, say.

For $\lambda = -2$, we solve $(A + 2I)\mathbf{x} = \mathbf{0}$, or $\begin{pmatrix} 3 & 3 \\ 4 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, i.e., $(x_1, x_2) = (-b, b)$ for $b \in \mathbb{R}$. So we take $\mathbf{v}_2 = (-1, 1)$, say.

So $P = \begin{pmatrix} 3 & -1 \\ 4 & 1 \end{pmatrix}$ will do. We can check that $P^{-1}AP = D$, or, what is simpler and equivalent, that $AP = PD$.

LECTURE 16

7 Polynomials

7.1 *Definition*

Let V be a vector space over F and $T : V \to V$ a linear mapping. Let $p(t) = a_0 + a_1 t + \ldots + a_m t^m$ be a polynomial with coefficients in F. Then we define
$p(T) = a_0 I + a_1 T + a_2 T^2 + \ldots + a_m T^m$, i.e.,
$p(T)\mathbf{v} = a_0 \mathbf{v} + a_1 T(\mathbf{v}) + a_2 T(T(\mathbf{v})) + \ldots + a_m T(T \ldots (\mathbf{v}))$.

Note that if T is represented by a matrix A with respect to a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, then T^k is represented by A^k (induction on k, using (4.13)), and $p(T)$ is represented by $p(A) = a_o I_n + a_1 A + \ldots + a_m A^m$.

7.2 *Definition*

Let V be an n-dimensional vector space over F, and $T : V \to V$ linear. Suppose that T is represented by an $n \times n$ matrix A with respect to some basis. Then the *characteristic polynomial* of T is $\chi(\lambda) = \det(\lambda I_n - A)$. This is independent of the choice of basis, since if $B = P^{-1}AP$ represents T with respect to another basis, then $\det(\lambda I_n - B) = \det(\lambda I_n - A)$ by (6.7).

N.B. Some people use $\det(A - \lambda I_n)$ as the definition of $\chi(\lambda)$. We accept either definition, as they only differ by a factor of $(-1)^n$, and so have the same roots.

7.3 *The Cayley–Hamilton theorem*

Let A be a real or complex square matrix, and $\chi(\lambda)$ its characteristic polynomial. Then $\chi(A) = O$, the zero matrix.

Proof: To be discussed later. Note $\chi(A) = \det(\lambda I_n - A)$, but we can't just substitute $\lambda = A$.

Example

Take $A = \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix}$, as in (6.15). Then $\chi(\lambda) = \lambda^2 - 3\lambda - 10 = (\lambda - 5)(\lambda + 2)$, and

$$(A - 5I)(A + 2I) = \begin{pmatrix} -4 & 3 \\ 4 & -3 \end{pmatrix} \begin{pmatrix} 3 & 3 \\ 4 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Similarly, if $T : V \to V$ is a linear mapping on an n-dimensional vector space, then $\chi(T) = 0$, because $\chi(\lambda)$ is defined in terms of a matrix representing T.

So we know that there are polynomials which "kill" a matrix A. It is important to find the simplest one.

7.4 *Definition*

The *minimum* or *minimal* polynomial of a square matrix A is the monic polynomial of least degree such that $\mu(A) = O$. ("Monic" means that the leading coefficient is 1.) We write it μ or μ_A.

Example. $A = \begin{pmatrix} 4 & 2 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}$ has $\chi_A(\lambda) = \det(\lambda I - A) = (\lambda - 4)^3$ (check).

So $(A-4I)^3 = O$. But in fact $(A-4I)^2 = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, and the minimum polynomial is $\mu_A(\lambda) = (\lambda - 4)^2$.

Now, given the characteristic polynomial, we can show that there are only a small number of possibilities to test for the minimum polynomial.

7.5 *Theorem*

Let A be a square matrix; then:
(i) Every eigenvalue of A is a root of the minimum polynomial;
(ii) The minimum polynomial divides the characteristic polynomial exactly.

Proof: (i) Let $A\mathbf{x} = \lambda\mathbf{x}$ with $\mathbf{x} \neq \mathbf{0}$. Then $A^k\mathbf{x} = \lambda^k\mathbf{x}$ for $k = 1, 2, 3, \ldots$, and by taking linear combinations we get that $p(A)\mathbf{x} = p(\lambda)\mathbf{x}$ for each polynomial p.

Now put $p = \mu$, and we get $\mu(A)\mathbf{x} = \mu(\lambda)\mathbf{x}$, and $\mu(A)\mathbf{x} = \mathbf{0}$. Since $\mathbf{x} \neq \mathbf{0}$, we have $\mu(\lambda) = 0$.

(ii) Let μ be the minimum and χ the characteristic polynomial. By long division we can write $\chi = \mu q + r$ for polynomials q (quotient) and r (remainder), with $\deg r < \deg \mu$.

Now

$$\begin{array}{ccccc} \chi(A) & = & \mu(A)q(A) & + & r(A) \\ O & & O & & \\ CH\ (7.3) & & \text{Defn. of } \mu & & \end{array}$$

so $r(A) = O$. Since $\deg r < \deg \mu$ and μ is the minimum polynomial, we have $r \equiv 0$, i.e., μ divides χ exactly.

□

LECTURE 17

Example.

Let $A = \begin{pmatrix} 3 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 6 \end{pmatrix}$. Calculate μ_A and χ_A.

Then

$$\det(\lambda I - A) = \begin{vmatrix} \lambda-3 & -1 & 0 & 0 \\ 0 & \lambda-3 & 0 & 0 \\ 0 & 0 & \lambda-6 & 0 \\ 0 & 0 & 0 & \lambda-6 \end{vmatrix} = (\lambda-3)^2(\lambda-6)^2.$$

So the eigenvalues are 3 and 6 only. We have $\chi(t) = (t-3)^2(t-6)^2$, and $\mu(t)$ has to divide it; also, both $(t-3)$ and $(t-6)$ must be factors of $\mu(t)$.

The only possibilities are therefore:

$$\begin{array}{cc} (t-3)(t-6) & \text{degree } 2, \\ (t-3)^2(t-6) & \text{degree } 3, \\ (t-3)(t-6)^2 & \text{degree } 3, \\ (t-3)^2(t-6)^2 & \text{degree } 4. \end{array}$$

We try them in turn. So

$$(A-3I)(A-6I) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} -3 & 1 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \neq O.$$

Next try

$$(A-3I)^2(A-6I) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = O,$$

so that $\mu(t) = (t-3)^2(t-6)$.
We can check that $(A-3I)(A-6I)^2 \neq O$.
Of course $(A-3I)^2(A-6I)^2 = O$, by Cayley–Hamilton.

7.6 *Proposition*

Similar matrices have the same minimum polynomial. Hence if A is diagonalisable, then $\mu_A(t)$ has no repeated roots.

Proof: If $B = P^{-1}AP$, then $B^n = (P^{-1}AP)(P^{-1}AP)\ldots(P^{-1}AP) = P^{-1}A^nP$, and so for any polynomial p we have $p(B) = P^{-1}p(A)P$.

Hence $p(B) = 0 \iff p(A) = 0$. So $\mu_B(t) = \mu_A(t)$.

Now, if B is a diagonal matrix, say $\begin{pmatrix} \lambda_1 & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & \ddots & 0 & \dots & \dots & \dots & 0 \\ 0 & 0 & \lambda_1 & 0 & \dots & \dots & 0 \\ 0 & \dots & 0 & \ddots & 0 & \dots & 0 \\ 0 & \dots & \dots & 0 & \lambda_k & 0 & 0 \\ 0 & \dots & \dots & \dots & 0 & \ddots & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 & \lambda_k \end{pmatrix}$ (with possibly repeated diagonal entries), then $(B-\lambda_1 I)(B-\lambda_2 I)\dots(B-\lambda_k)I = O$, so $\mu_B(t)$ has no repeated roots since it must divide the polynomial $(t-\lambda_1)(t-\lambda_2)\dots(t-\lambda_k)$. But any diagonalisable A is similar to a matrix B of this form, so μ_A has no repeated roots.

8 The Jordan canonical form

From (6.9), we know that a matrix is diagonalisable if and only if there is a basis consisting of its eigenvectors. What can we do if this is not the case?

8.1 *Definition*

A *Jordan block matrix* is a square matrix of the form $J_\lambda = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \ddots & \lambda & 1 \\ 0 & \dots & \dots & 0 & \lambda \end{pmatrix}$, with λ on the diagonal and 1 just above the diagonal (for some fixed scalar λ). There is also the trivial 1×1 case.

For example, the following are Jordan block matrices:

$$(4), \quad \begin{pmatrix} 4 & 1 \\ 0 & 4 \end{pmatrix}, \quad \begin{pmatrix} 4 & 1 & 0 \\ 0 & 4 & 1 \\ 0 & 0 & 4 \end{pmatrix}, \quad \begin{pmatrix} 4 & 1 & 0 & 0 \\ 0 & 4 & 1 & 0 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 4 \end{pmatrix}.$$

8.2 *Proposition*

Suppose that V is an n-dimensional vector space, and T a linear transformation on V, represented by a Jordan block matrix A with respect to some basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$.

Then $\chi(t) = \det(tI - A) = (t - \lambda)^n$, and we also have

$$\begin{aligned} T\mathbf{v}_1 &= \lambda\mathbf{v}_1, \\ T\mathbf{v}_2 &= \mathbf{v}_1 + \lambda\mathbf{v}_2, \\ T\mathbf{v}_3 &= \mathbf{v}_2 + \lambda\mathbf{v}_3, \\ \dots & \quad \dots \\ T\mathbf{v}_n &= \mathbf{v}_{n-1} + \lambda\mathbf{v}_n. \end{aligned}$$

So, if we define $\mathbf{v}_0 = \mathbf{0}$ we have $(T - \lambda I)\mathbf{v}_k = \mathbf{v}_{k-1}$ for $k = 1, 2, \dots, n$. Hence $(T - \lambda I)^k \mathbf{v}_k = \mathbf{0}$.

Proof: To get $\chi(t)$ we expand $\begin{vmatrix} t-\lambda & -1 & 0 & \dots & 0 \\ 0 & t-\lambda & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \ddots & t-\lambda & -1 \\ 0 & \dots & \dots & 0 & t-\lambda \end{vmatrix}$ about the first column, then continue. To work out what $T\mathbf{v}_k$ is, just look at the kth column of A. The rest is clear.

□

LECTURE 18

8.3 *Definition*

Let V be a vector space over F and $T : V \to V$ a linear mapping. Let $\lambda \in F$ be an eigenvalue of T. The non-zero vector $\mathbf{v} \in V$ is said to be a *generalized eigenvector* of T corresponding to λ if $(T - \lambda I)^k \mathbf{v} = \mathbf{0}$ for some $k \geq 1$.
Similarly for $n \times n$ matrices A, a column vector $\mathbf{x}$ is a generalized eigenvector if $(A - \lambda I)^k \mathbf{x} = \mathbf{0}$ for some $k \geq 1$.

Clearly every eigenvector is a generalized eigenvector (take $k = 1$).

Example.
For $A = \begin{pmatrix} 4 & 1 \\ 0 & 4 \end{pmatrix}$, we have $\mathbf{v}_1 = (1, 0)$ an eigenvector with eigenvalue $\lambda = 4$, since $\begin{pmatrix} 4 & 1 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$. Now $\mathbf{v}_2 = (0, 1)$ is not an eigenvector, since $\begin{pmatrix} 4 & 1 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \end{pmatrix} = \mathbf{v}_1 + 4\mathbf{v}_2$. So $(A - 4I)\mathbf{v}_2 = \mathbf{v}_1$, and then $(A - 4I)^2 \mathbf{v}_2 = (A - 4I)\mathbf{v}_1 = 0$; so $\mathbf{v}_2$ is a generalized eigenvector.

8.4 *Definition*

A square matrix A is said to be in *Jordan canonical form* (or *Jordan normal form*) if it consists of Jordan block matrices strung out on the diagonal, with zeroes elsewhere. The diagonal entries are the eigenvalues of A.

8.5 *Examples*

Any Jordan block matrix is in JCF. So are:

$$A = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \qquad B = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}, \qquad \text{and} \qquad C = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}.$$

here:
A has 3 blocks all of size 1×1;
B has 1 block of size 1×1 and 1 of size 2×2;
C has 1 block size 3×3 and 1 block size 2×2.

8.6 *Theorem*

Let A be an $n \times n$ matrix with entries in $\mathbb{C}$. Then A is similar to a matrix in Jordan canonical form, unique up to re-ordering the blocks. If A is an $n \times n$ matrix with real entries, then A is similar to a **real** matrix in JCF (i.e., $B = P^{-1}AP$ with B real) if and only if all the roots of the characteristic equation are real.

Proof: Omitted, but the "only if" follows from the fact that if $B = \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_N \end{pmatrix}$, is in JCF and each B_i is a Jordan block of size m_i with diagonal elements λ_i, then

$$\chi_B(\lambda) = \chi_{B_1}(\lambda) \dots \chi_{B_N}(\lambda) = (\lambda - \lambda_1)^{m_1} \dots (\lambda - \lambda_N)^{m_N},$$

so if all the blocks are real then χ has only real roots.

8.7 *Facts which help us find the JCF of a matrix A*

1. For each λ the power of $(t - \lambda)$ in $\chi(t)$ is the total size of Jordan blocks using λ. See (8.6).

2. The number of λ-blocks is the dimension of the eigenspace $\ker(A - \lambda I)$. For each block gives one new eigenvector, by (8.2).

3. The biggest blocksize for a λ-block is the power of $(t - \lambda)$ in the minimum polynomial. For we need to take $(J_\lambda - \lambda I)^n$ to kill a Jordan block J_λ of size n, by (8.2).

8.8 *Example*

Take $A = \begin{pmatrix} 5 & 0 & -1 \\ 2 & 3 & -1 \\ 4 & 0 & 1 \end{pmatrix}$. Find the Jordan canonical form of A.

N.B. We will do this at the very end of the course for everything up to 4×4 matrices.

We always start by finding

$$\begin{aligned} \chi(\lambda) &= \det(\lambda I - A) = \begin{vmatrix} \lambda - 5 & 0 & 1 \\ -2 & \lambda - 3 & 1 \\ -4 & 0 & \lambda - 1 \end{vmatrix} \\ &= (\lambda - 3) \begin{vmatrix} \lambda - 5 & 1 \\ -4 & \lambda - 1 \end{vmatrix} \\ &= (\lambda - 3)(\lambda^2 - 6\lambda + 5 + 4) = (\lambda - 3)^3. \end{aligned}$$

So 3 is the only eigenvalue.

We can solve $(A - 3I) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ or

$\begin{pmatrix} 2 & 0 & -1 \\ 2 & 0 & -1 \\ 4 & 0 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$. This is just $2x = z$ and so

$$\ker(A - 3I) = \{(a, b, 2a) : a, b \in \mathbb{R}\},$$

a 2-dimensional space of eigenvectors (with $\mathbf{0}$).

This tells us that there are 2 blocks, i.e., one size 2 and one size 1.

Alternatively, we can check that $(A - 3I) \neq O$, but

$(A - 3I)^2 = \begin{pmatrix} 2 & 0 & -1 \\ 2 & 0 & -1 \\ 4 & 0 & -2 \end{pmatrix}^2 = O.$

Thus the minimum polynomial of A is $(t - 3)^2$, and the largest blocksize is size 2, so it must be $2 + 1$ again.

The Jordan form of A is $\begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$, with one Jordan block of size 2×2 and one of size 1×1.

LECTURE 19

8.9 *The Cayley–Hamilton theorem via Jordan canonical matrices*

Let A be a real or complex matrix and $\chi(t)$ the characteristic polynomial. Then $\chi(A) = O$.

Proof: Suppose that M is a Jordan canonical matrix. If $M = \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_m \end{pmatrix}$, with $B_1, \ldots, B_m$ Jordan blocks, then $M^2 = \begin{pmatrix} B_1^2 & & 0 \\ & \ddots & \\ 0 & & B_m^2 \end{pmatrix}$, and similarly for higher powers of M. So $\chi(M) = \begin{pmatrix} \chi(B_1) & & 0 \\ & \ddots & \\ 0 & & \chi(B_m) \end{pmatrix}$ and we need to show that $\chi(B_i) = 0$ for each i.

Now suppose that $B = \begin{pmatrix} \lambda & 1 & 0 & \ldots & 0 \\ 0 & \lambda & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & \ddots & \lambda & 1 \\ 0 & \ldots & \ldots & 0 & \lambda \end{pmatrix}$ is a Jordan block of size k. Then $(B - \lambda I)^k = O$ as in (8.2).
Thus $\chi(B) = 0$ for each of the blocks B making up A in Jordan form (since in each case $(t - \lambda)^k$ divides $\chi(t)$).

Now $\chi(M) = \begin{pmatrix} \chi(B_1) & & 0 \\ & \ddots & \\ 0 & & \chi(B_m) \end{pmatrix} = O.$

Finally, any matrix A is similar to a matrix M in Jordan canonical form with the same characteristic polynomial, by (6.7), i.e., $M = P^{-1}AP$, and
$\chi(A) = P^{-1}\chi(M)P = O.$

□

8.10 *Theorem*

If no eigenvalue λ appears with multiplicity greater than 3 in the characteristic equation for A, then we can write down the JCF knowing just the characteristic and minimum polynomials ($\chi(t)$ and $\mu(t)$).

Proof: **In general the power of $(t-\lambda)$ in the characteristic polynomial is the total number of diagonal entries λ we get in the JCF; the power**

of $(t-\lambda)$ in the minimum polynomial is the size of the largest Jordan block associated with λ.

If the multiplicity of $(t-\lambda)$ in $\chi(t)$ is 1, then there is just one block, size 1. The power of $(t-\lambda)$ in both $\chi(t)$ and $\mu(t)$ is 1.

If the multiplicity of $(t-\lambda)$ in $\chi(t)$ is 2, then we have either one block, size 2, or two blocks size 1, that is, looking just at the part corresponding to this λ, we have

$$\begin{array}{ccc} \text{block sizes} & 2 & 1+1 \\ \text{matrices} & \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} & \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \\ \chi & (t-\lambda)^2 & (t-\lambda)^2 \\ \mu & (t-\lambda)^2 & (t-\lambda). \end{array}$$

For multiplicity 3, there may be one block size 3, a $2+1$ or a $1+1+1$.

$$\begin{array}{cccc} \text{block sizes} & 3 & 2+1 & 1+1+1 \\ \text{matrices} & \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} & \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} & \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} \\ \chi & (t-\lambda)^3 & (t-\lambda)^3 & (t-\lambda)^3 \\ \mu & (t-\lambda)^3 & (t-\lambda)^2 & (t-\lambda). \end{array}$$

□

8.11 *Example*

A matrix A has characteristic polynomial $\chi(t) = (t-1)^3(t-2)^3(t-3)^2$ and minimal polynomial $\mu(t) = (t-1)^3(t-2)(t-3)^2$. Find a matrix in Jordan form similar to A.

Solution: For $\lambda = 1$, there is one block size 3; for $\lambda = 2$, there are 3 blocks size 1; for $\lambda = 3$ there is one block size 2.

Remark: If the eigenvalue has multiplicity 4, the possibilities are now 4, $3+1$, $2+2$, $2+1+1$, $1+1+1+1$. For both $2+2$ and $2+1+1$ we have $\chi(t) = (t-\lambda)^4$ and $\mu(t) = (t-\lambda)^2$. So χ and μ alone don't help us distinguish between

$$\begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \lambda \end{pmatrix},$$

since both have the largest block of size 2.

However, it is still possible to determine quickly which case we are in, as in the first case there are 2 blocks, and the eigenspace $\ker(A - \lambda I)$ has dimension 2; in the other case, 3 blocks, and it has dimension 3.

LECTURE 20

8.12 *Worked example (from the 2003 paper)*

You are given that the matrix $A = \begin{pmatrix} 7 & 1 & 1 & 1 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 6 & -2 \\ -1 & 1 & -1 & 7 \end{pmatrix}$ has characteristic polynomial $\chi_A(t) = (t-6)^2(t-8)^2$. Find its Jordan canonical form and its minimum polynomial.

Solution: We see that the eigenvalues are 6 and 8. Let's go for the minimum polynomial μ first. This has roots 6 and 8, and divides χ. Since it is monic, it is therefore one of:
$(t-6)(t-8)$, $(t-6)^2(t-8)$, $(t-6)(t-8)^2$, or $(t-6)^2(t-8)^2$.

We calculate

$$(A-6I)(A-8I) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & -2 \\ -1 & 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & -2 \\ -1 & 1 & -1 & -1 \end{pmatrix} = \begin{pmatrix} -2 & 2 & -2 & -2 \\ 0 & 0 & 0 & 0 \\ 2 & -2 & 2 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix} \neq O,$$

which eliminates the first possibility. Next,

$$(A-6I)^2(A-8I) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & -2 \\ -1 & 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} -2 & 2 & -2 & -2 \\ 0 & 0 & 0 & 0 \\ 2 & -2 & 2 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = O,$$

so the minimum polynomial is $(t-6)^2(t-8)$. (If it hadn't been zero, we would then have tried $(t-6)(t-8)^2$.)

Each eigenvalue has multiplicity 2, so we have to work out whether the blocks are 2 or $1+1$. Since we needed $(t-6)^2$, this block is size 2, and since we only needed $(t-8)^1$ this has two blocks of size 1. The Jordan form is therefore:

$$B = \begin{pmatrix} 6 & 1 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix}.$$

THE END