M62

N 2735

LINEAR ALGEBRA

PROF. O'NEILL
2-176

M 62

Definition of Vector Space: Vector Space consists of:

I A collection of things (called vectors) together with a field of objects called scalars (in this course, real & complex numbers)

II The operations:

(1) addition of vectors: given two vectors a & b, there is a vector $a+b$ called the sum of a & b.

(2) scalar multiplication: given a vector b & a scalar λ, there is a unique vector λb

III The operations satisfy the following axioms:

(1) Vector addition:

commutative: $a+b = b+a$

associative: $(a+b)+c = a+(b+c)$

zero vector: there exists a vector $\vec{0}$ such that $b+\vec{0} = b$

inverse: for each vector b, there is a vector $(-b)$ such that $b+(-b)=\vec{0}$

(2) Scalar multiplication

associative: $\lambda(\mu a) = (\lambda\mu)a$

identity: $1a = a$ for any a

(3) Distributive

$\lambda(a+b) = \lambda a + \lambda b$

$(\lambda+\mu)a = \lambda a + \mu a$

a, b, c any vectors
λ, μ any scalars

Birkhoff & Mc Lane
Modern Algebra
Chapter on vector spaces.

Thrall & Tornheim
Vector Spaces & Matrices

Paul Halmos
Finite Dimensional Vector Spaces

Lecture 2/14

There is an intuitive idea of what a line, plane, & 3-space are. We believe ~~in~~ we can draw arrows (called "vectors"), add by parallelogram rule, & scalar multiply. The result is a vector space - if you trust intuition. But we need a mathematical definition of 3-space in terms of simple standard mathematical objects. Motivation for this definition is the idea of coordinates. These assign to each point of 3-space an ordered triple of numbers (x_1, x_2, x_3). This assignment is one-to-one.

Now we take a big step - abandon the intuitive idea of 3-space and take the set R_3 of all triples of numbers as the object of study. We continue to use our intuition about "physical space" to guide us in setting up a definition of things in R_3.

In particular, we want to give precise mathematical definitions of objects in R_3 (vectors) which correspond to the arrows that we think of ~~as~~ as being in "physical space." There are 3 major choices of how to do this identical in the way they behave but different in what they are (isomorphism):

(1) Very complicated - physicists' way of handling forces on a on a body when interested only in the motion of c.g., not rotation

(2) Complicated - same as (1) except pt. of application matters.

(3) Simple - mathematicians' choice.

To describe these choices, we need a definition: the oriented line segment (or "arrow") in R_3 from (x_1, x_2, x_3) to (y_1, y_2, y_3) is the set of all triples in R_3 which can be written in the form $((1-t)x_1 + ty_1, (1-t)x_2 + ty_2, (1-t)x_3 + ty_3)$ for some number t, $0 \le t \le 1$

I. "Family" definition of vector in R_3

Consider all oriented line segments in R_3. These break up into families ~~a~~ according to the following rule: The oriented line segment from (x_1, x_2, x_3) to (y_1, y_2, y_3) and the " " " " $(\bar{x}_1, \bar{x}_2, \bar{x}_3)$ to $(\bar{y}_1, \bar{y}_2, \bar{y}_3)$ are in the family if & only if

$$y_i - x_i = \bar{y}_i - \bar{x}_i \qquad (i = 1, 2, 3)$$

Intuitively, this means that the two arrows have the same length and direction, although they may have different points of application. Every oriented line segment is in one and only one family. <u>These families are called vectors</u>. A <u>family</u> is a vector.

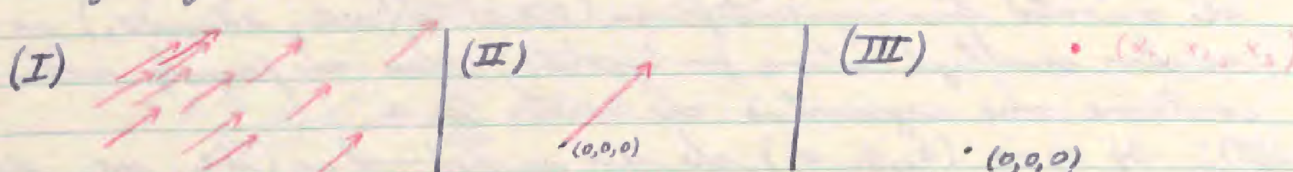
II. <u>Oriented line segment</u> definition of vector

A vector is an oriented line segment with its point of application ("tail") at the origin [(0,0,0)].

III <u>Mathematical</u> definition

A vector is an ordered triple of numbers, i.e., a point in R_3.

Drawings of <u>one</u> (1) vector according to each definition:



These three definitions are different but they behave the same.

There is a natural one-to-one correspondence between the set of all vectors and all vectors of type 2 and all vectors of type 3. The correspondence is set by noting that an ordered triple of numbers (x_1, x_2, x_3) determines a vector of each of the three types: (II) Determines the oriented line segment from $(0,0,0)$ to (x_1, x_2, x_3); (I) Determines the family containing type (II) vector; and (III) just the triple (x_1, x_2, x_3). Since the triple immediately determines each of the three types of vectors, we need consider <u>only</u> the ordered triple of numbers; the rest is superfluous.

Definition of vector addition in each of the 3 cases.

<u>Type (II)</u>: If a is the oriented line segment from $(0,0,0)$ to (a_1, a_2, a_3)
and b " " " " " " " (b_1, b_2, b_3),
then $(a+b)$ is defined as the oriented line segment from $(0,0,0)$ to $(a_1+b_1, a_2+b_2, a_3+b_3)$. This definition is justified since it is modelled on the parallelogram rule, and since $(a+b) - a = b$.

Type (I): If a & b ~~and~~ are families of oriented line segments, in each pick an oriented line segment that has its initial point at (0,0,0). Add these two as in type (II) definition of vector addition; then the family which contains the sum is the vector $(a+b)$.

Type (III): If a is (a_1, a_2, a_3) and b is (b_1, b_2, b_3) then $(a+b)$ is defined as $(a_1+b_1, a_2+b_2, a_3+b_3)$.

Definition of scalar multiplication for each definition of a vector.

Type (II): If a is the oriented line segment from the origin to (a_1, a_2, a_3), and u is any scalar, then the vector ua is defined as the oriented line segment from (0,0,0) to (ua_1, ua_2, ua_3).

Type (I): If a is a family of oriented line segments, ~~of length~~ pick one one oriented line segment that has its initial point at (0,0,0) and hence, its terminal point at (a_1, a_2, a_3). Form the oriented line segment from (0,0,0) to (ua_1, ua_2, ua_3) as in scalar multiplication for type (II); then the family which contains this segment is the vector ua.

Type (III): If a is (a_1, a_2, a_3) the ua is defined as (ua_1, ua_2, ua_3).

The precise meaning of the statement that the three types of vectors in R_3 behave the same, say consider (I) and (III), is this: the natural one-to-one correspondence between the collection of all family type vectors and the collection (R_3 itself) of all points of R_3 possess vector addition and scalar multiplication; that is, if $(a_1, a_2, a_3) \leftrightarrow$ family A and if $(b_1, b_2, b_2) \leftrightarrow$ family B, then $(a_1, a_2, a_3) + (b_1, b_2, b_3) \leftrightarrow A+B$.

Thus, the operations of vector addition and scalar multiplication are essentially the same for all three types and since vector properties are all based on vector addition and scalar multiplication, the three are essentially the same in all "vector properties". So we can drop definitions (I) and (II) and use only definition (III).

Examples of vector spaces

(1) R_3 is a vector space. (with operations)

(2) R_n (n is any positive integer) is the set of all n-tuples $(x_1, x_2, \cdots, x_n)$ of real numbers; vector addition defined - to add two n-tuples, add their corresponding components; scalar multiplication defined - to multiply a vector by a scalar, multiply each of its components by the scalar.

R_n with the two defined operations is a vector space.

(3) Let V = set of all power series ~~on~~ which converge everywhere on the real line; the scalars are the real numbers.

Vector addition defined →

$$\sum_{i=0}^{\infty} a_i x^i + \sum_{i=0}^{\infty} b_i x^i = \sum_{i=0}^{\infty} (a_i + b_i) x^i$$

Scalar multiplication defined →

$$\lambda \sum_{i=0}^{\infty} a_i x^i = \sum_{i=0}^{\infty} \lambda a_i x^i$$

V with the two defined operations is a vector space.

(4) Let V be the set of all continuous (real valued) functions on the interval $0 \le t \le 1$; scalars are the real numbers.

Vector addition defined →

Usual addition of functions: $f(t) + g(t) = (f+g)(t)$

Scalar multiplication defined →

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

In order for anything to be a vector space, it must satisfy the definition of a vector space; i.e., the axioms set forth on the first page of these notes.

Span and linear independence.

Definition of "span".

A finite number of vectors $a_1, a_2, \ldots, a_s$ span a vector space V provided every vector in V can be written as a linear combination of the a_i's; i.e., $v = \sum_{i=1}^{s} \lambda_i a_i$

Roughly, this means the vectors a_i "control" V.

Definition of linear independence:

A finite number of vectors $b_1, b_2, \ldots, b_\ell$ in V are linearly independent provided the only linear combination of the b_j's which is zero is the trivial one; i.e., if $\sum_{j=1}^{\infty} \lambda_j b_j = 0$, then $\lambda_1 = \lambda_2 = \lambda_3 = \ldots = \lambda_\ell = 0$

Roughly, this means that each b_j points into a different dimension.

Proof by induction:

Fact from logic: Let $P(n)$ be a proposition (not necessarily true) which depends on a non-negative integer n, e.g., $P(n) \Leftrightarrow (n+1)^2 = n^2+2n+1$ or $P(n) \Leftrightarrow n^n = (n!)^2$; then if $P(0)$ is true and if the truth of $P(n)$ implies the truth of $P(n+1)$ for all n, then $P(n)$ is true for all integers.

Replacement Theorem

If vectors $a_1, a_2, \ldots, a_s$ span V and vectors $b_1, b_2, \ldots, b_\ell$ in V are linearly independent, then $\ell \leq s$ and (possibly after reordering), $b_1, b_2, \ldots, b_\ell, a_{\ell+1}, \ldots, a_s$ span V.

Proof: By induction on the non-negative integer ℓ.

If $\ell = 0$, then the theorem is trivially true: $0 \leq s$ and $a_1, a_2, \ldots, a_s$ span V (by definition).

Suppose the theorem is true for $\ell - 1$; so we can apply the theorem to:

$a_1, a_2, \ldots, a_s$ span V & $b_1, b_2, \ldots, b_{l-1}$ are linearly independent,

to conclude that $(l-1) \le s$ & $b_1, b_2, \ldots, b_{l-1}, a_l, a_{l+1}, \ldots, a_s$ span V.

Thus since that set of vectors spans V, we can write

$$b_l = \sum_{i=1}^{l-1} \lambda_i b_i + \sum_{i=l}^{s} \lambda_i a_i$$

Now the numbers $\lambda_l, \ldots, \lambda_s$ exist and are not all zero for otherwise, we would have a non-trivial linear relation among the b_i's; but that is impossible since they are linearly independent. → Thus, $(l-1) < s$, so $l \le s$

By (possibly) relabelling, we can assume $\lambda_l \neq 0$. Then we get

$$b_l = \sum_{i=1}^{l-1} \lambda_i b_i + \lambda_l a_l + \sum_{l+1}^{s} \lambda_i a_i \qquad \text{with } \lambda_l \neq 0$$

Thus,

$$a_l = \frac{1}{\lambda_l} b_l - \sum_{i=1}^{l-1} \frac{\lambda_i}{\lambda_l} b_i - \sum_{i=l+1}^{s} \frac{\lambda_i}{\lambda_l} a_i$$

Now, let v be any vector. Can express v as a linear combination of $b_1, \ldots, b_{l-1}, a_l, \ldots, a_s$. Then substituting the above expression for a_l, gives v as a linear combination of $b_1, \ldots, b_l, a_{l+1}, \ldots, a_s$. Hence, $b_1, \ldots, b_l, a_{l+1}, \ldots, a_s$ span V. Q.E.D.

Trivial properties of vectors (immediately obtainable from definition)

1. Uniqueness of the zero vector
2. Uniqueness of the negative vectors.
3. If z is a vector such that for some one vector v_0, we have $v_0 + z = v_0$, then $v + z = v$ for every vector v, i.e., $z = \vec{0}$
4. $-\vec{0} = \vec{0}$
5. $0v = \vec{0}$ for all v
6. $\lambda \vec{0} = \vec{0}$
7. $(-1)x = -x$ for x any vector
8. If $\lambda v = \vec{0}$, then either $\lambda = 0$, or $v = \vec{0}$

A non-empty set of integers which is bounded above contains a largest integer: axiom of logic.

<u>Definition</u>: A vector space V is finite dimensional provided it can be spanned by a finite set of vectors.

For a finite dimensional V, let $\mathcal{P}$ be the set of all integers j with the property that there are j linearly independent vectors in V.

Then, (1) $\mathcal{P}$ is not empty [V always contains $\vec{O}$ so $\mathcal{P}$ contains at least one integer.]
(2) $\mathcal{P}$ is bounded above.

Now, the dimension p of V is defined to be the largest integer in $\mathcal{P}$.

<u>Definition</u>: A set of vectors in V which spans V and is linearly independent is called a <u>basis</u> for V.

[<u>Span + Linear Independence = Basis → Dimension</u>]

<u>Theorem</u>: Every finite dimensional vector space has a basis. Every basis of V consists of (exactly) p vectors where $p = \dim V$.

<u>Proof:</u>
Since V is finite dimensional [i.e., there is a maximal linearly independent set of vectors], we know that V contains a maximal linearly independent $a_1, \dots, a_p$ where p is (by definition) $\dim V$.

Claim $a_1, \dots, a_p$ spans V.: Let v be any vector in V. Consider $a_1, \dots, a_p, v$. This set is linearly dependent because p is the maximal number of ~~vectors~~ linearly independent vectors. So,

$$\sum_{i=1}^{p} \lambda_i a_i + \lambda v = \vec{O} \quad \text{with} \begin{cases} \lambda \neq 0 \text{ since } a_1, \dots, a_p \text{ are lin. ind.} \\ \text{not all } \lambda_i\text{'s} + \lambda = 0 \end{cases}$$

So this is solvable for $v = -\sum \frac{\lambda_i}{\lambda} a_i$.

By definition of basis plus the replacement theorem, any two bases have the same number of vectors (P).

$\left[\begin{array}{l} l \text{ ind vectors \& possible } s \text{ spanning vectors} \div \; s \leq l \text{ by def of } \cancel{\text{basis}} \text{ dimension.} \\ l \leq s \text{ by replacement theorem \& } \therefore \; s = l \end{array}\right]$

Example: R_n has dimension n

Proof:

By theorem stating every basis of V contains $p = \dim V$ vectors, it is sufficient to show $e_1, \dots, e_n$ is a basis where $e_i = (0, 0, \dots, 1, \dots, 0)$ where 1 is the i^{th} term.

$e_1, \dots, e_n$ span since $(\alpha_1, \dots, \alpha_n) = \sum_{i=1}^{n} \alpha_i e_i$

If $\sum \lambda_i e_i = \vec{0}$, then $(\lambda_1, \dots, \lambda_n) = (0, \dots, 0)$; i.e., $\lambda_1 = \lambda_2 = \dots = \lambda_n = 0$

Hence $e_1, \dots, e_n$ span R_n & are linearly independent & $\therefore$ constitute a basis.

SUB-SPACES

Definition: A set W of vectors in V is called a sub-space of V provided that if a & b are in W, then $a + b$ & λa are in W.

Note that if the scalar multiplication and vector addition operations of V are applied to W, then W becomes a vector space.

If V & W are subspaces of a vector space L,

let $\begin{cases} V \cap W \text{ be the set of vectors in both } V \text{ \& } W. \\ V + W \text{ be the set of all vectors } a \text{ in } L \text{ which can be written} \\ \quad a = v + w \text{ where } v \text{ is in } V \text{ and } w \text{ is in } W. \end{cases}$

Theorem: $V \cap W$ and $V + W$ are also sub-spaces.

Example: R_3: $V =$ all vectors $(\alpha_1, 0, 0)$
$W =$ all vectors $(0, \alpha_2, \alpha_3)$

then $\begin{cases} V \cap W \text{ is } \vec{0} \text{ alone } (0,0,0) \\ V + W \text{ is } R_3 \end{cases}$

<u>Theorem</u>: $\dim(V+W) + \dim(V \cap W) = \dim V + \dim W$

Let $a_1, \dots, a_v$ be a basis for V such that $a_1, \dots, a_\alpha$ is in $V \cap W$
Let $b_1, \dots, b_w$ be a basis for W " " $b_1, \dots, b_\alpha$ " " "

Further, let $b_1, \dots, b_\alpha$ & $a_1, \dots, a_\alpha$ each be a basis for $V \cap W$ ~~and in $V \cap W$~~

Now $\alpha \le v$ from def of $V \cap W$ [$V \cap W$ in V]

Now, $a_1, \dots, a_v, b_1, \dots, b_w$ span $V+W$ since $V+W$ consists of vectors ~~in with~~ $a + b = \sum \lambda_i a_i + \sum \mu_i b_i$.

Also, $a_{\alpha+1}, a_{\alpha+2}, \dots, a_v, b_1, \dots, b_w$ span $V+W$ since those ~~vectors a in V which require $a_1, \dots, a_\alpha$ to be a linear combination of the a_i's~~ vectors $a_1, \dots, a_\alpha$ are also in W & hence can be written ~~$a = \Sigma$~~

$a_i = \sum_{j=1}^{\alpha} \lambda_j b_j$ where $1 \le i \le \alpha$;

~~Now assume $a_{\alpha+1}, \dots, a_v, b_1, \dots, b_w$ linearly dependent.~~
~~Then $\sum_{\alpha+1}^{v} \lambda_i a_i + \sum_{1}^{w} \mu_i b_i$ with some $\neq 0$~~

~~Now let $P = \dim(V+W)$ & let $\beta = w + (v - \alpha)$~~
~~$P \neq \beta$~~
~~$\alpha + P \neq w + v$~~

$\Rightarrow$ some $\lambda \neq 0$ & some $\mu \neq 0$ since a's are lin ind as are b's.

$\Rightarrow \sum \lambda_i a_i = \sum \mu_i b_i$

But none of these a's are in W and no b is in V

hence can't write this & $\therefore$, $a_{\alpha+1}, \dots, a_v, b_1, \dots, b_w$ are linearly independent & since they span, form a basis of $V+W$ & hence $w + (v - \alpha) = \dim(V+W)$

$\dim(V+W) + \dim(V \cap W) = [w + (v - \alpha)] + \alpha = w + v = \dim V + \dim W$

QED

Theorem 2: Any set of more than p linear combinations of the vectors $a_1, \dots, a_p$ in L is linearly dependent.

Proof:

Consider the set S of all linear combinations of $a_1, \dots, a_p$. Then S is a subspace of L spanned by $a_1, \dots, a_p$. If $x_1, \dots, x_q$ $(q > p)$ are a linear combination of the a's, then the x's are in S. So can deal only with S. Now, the x's could not be independent for since $q > p$, that would contradict the replacement theorem.

LINEAR TRANSFORMATIONS OF VECTOR SPACES

Definition 1: A function from a set A to a set B is a rule f which assigns to each element a of A a unique element f(a) of B.
Notation: $f: A \to B$ read function f from A to B.

The functions used in calculus are mostly $f: R_1 \to R_1$ or $g: R_2 \to R_1$. In calculus, the function is usually given by an explicit formula, e.g., $f(x) = x^2 + 1$. This defines the function f which assigns to the number x the value x^2+1. Note that we use the letter f to represent the function (thought of as an entity) while f(x) is (in this case) a number.

[All functions (e.g., $\sqrt{x}$) can be & are to be refined to single valued (e.g., $+\sqrt{x}$ or $-\sqrt{x}$)]
Transformation = function = mapping.

Now we begin to study functions carrying one vector space into another. If such a function is to have significance in the study of vector spaces, it must be related to vector addition and scalar multiplication in an important way.

Definition: A function $T: L \to L'$ (where L & L' are vector spaces with the same scalars) is a linear transformation provided:

$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$ for all scalars α & β & vectors x & y of L.

or: 1) $T(\lambda x) = \lambda T(x)$

2) $T(x+y) = T(x) + T(y)$

Examples of linear transformations:

(1) Define $T: R_1 \to R_1$ by $T(\xi) = \alpha\xi$. Then T is linear.
But if we define $T: R_1 \to R_1$ by $T(\xi) = \alpha\xi + \beta$, then T is not linear. Hence, "linear" as used here rigorously means "linear homogeneous"; i.e., straight lines through the origin.

(2) Define $T: R_2 \to R_2$ by $T(\xi, \eta) = (\xi\cos\theta - \eta\sin\theta, \xi\sin\theta + \eta\cos\theta)$ where θ is a fixed number. This is a rotation of the plane through an angle θ.

(3) $T: R_2 \to R_1$ defined by $T(\xi, \eta) = (\xi, 0)$.
This is "projection on the x-axis."

Definition: Null space consists of all vectors x in L such that $T(x) = 0$ in L' where $T: L \to L'$.

The range of T consist of all the vectors x in L' that can be expressed as $x = T(v)$ where v is some vector in L for $T: L \to L'$.

Theorem: dim L = dim of Nullspace + dim of range
= nullity + rank
where "nullity" & "rank" are integers.

Proof: Let $a_1, \ldots, a_n$ be a basis for the null space of T (in L). Now can extend this to a basis $a_1, \ldots, a_n, a_{n+1}, \ldots, a_p$ for all L since any p linearly independent vectors form a basis of L (dim $L = p$). [Hence, n = nullity and p = dim L]. Applying T to this basis, gives $T(a_1) = \ldots = T(a_n) = 0$ because a_i $(i = 1, \ldots, n)$ is in the null space. Now, if $T(a_{n+1}), \ldots, T(a_p)$ is a basis for the range of T, this will prove the theorem since a basis by definition consists of q vectors (q = dim V).

(next page)

[Note that $T(a_{n+1}), \ldots, T(a_p)$ is a basis if it spans & the vectors are linearly independent.]

[Span?] Let v be a vector in the range of T, i.e., $v = T(u)$ for some vector u of L. Hence, $u = \sum_{i=1}^{p} \lambda_i a_i$; applying T,

$$v = T(u) = T(\Sigma \lambda_i a_i) = \Sigma \lambda_i (T(a_i)) = \sum_{i=1}^{n} \lambda_i (Ta_i) + \sum_{i=n+1}^{p} \lambda_i (Ta_i)$$

$v = \sum_{i=n+1}^{p} \lambda_i (Ta_i)$ since $\sum_{i=1}^{n} \lambda_i (Ta_i) = 0$ since $a_i \; (i=1, \ldots, n)$ is in the null space of L for $T: L \to L'$. Hence, $a_{n+1}, \ldots, a_p$ span the range of T.

[Linearly Independent?] Suppose $\sum_{i=n+1}^{p} \lambda_i (Ta_i) = 0$. Then must show that ~~$\lambda_{n+1} = \lambda_{n+2} = \ldots = \lambda_p = 0$ of necessity~~ this implies $\lambda_{n+1} = \ldots = \lambda_p = 0$.

$0 = \sum_{i=n+1}^{p} \lambda_i (Ta_i) = T(\sum_{i=n+1}^{p} \lambda_i a_i)$; consequently, $\sum_{n+1}^{p} \lambda_i a_i$ is the null space. Hence, it can be expressed as a linear combination of the vectors $a_1, \ldots, a_n$; i.e.,

$$\sum_{i=n+1}^{p} \lambda_i a_i = \sum_{i=1}^{n} \mu_i a_i \text{ \underline{\& hence},} \quad \sum_{i=n+1}^{p} \lambda_i a_i + \sum_{i=1}^{n} (-\mu_i) a_i = 0 .$$

But $a_1, \ldots, a_p$ is a basis of L and the vectors are therefore linearly independent; hence, all $\lambda_i = \mu_i = 0$ and $T(a_{n+1}), \ldots, T(a_p)$ is linearly independent.

Now, since $T(a_{n+1}), \ldots, T(a_p)$ span the range of T and are linearly independent, they must be a basis of the range of T. Hence, $\dim R = p-n$ & $\dim \text{Nullspace} = n$ so, $\dim R + \dim N = p-n+n = p = \dim L$. Q.E.D.

<u>Definition</u>: A linear transformation $T: L \to L'$ is <u>onto</u> provided the range of T is L' (i.e., every vector in L' can be expressed as $T(v)$ where v is some vector in L).

A linear ~~t~~ transformation is <u>one-to-one</u> provided different vectors of L have different values in L' when acted on by $T: L \to L'$ (i.e., $T(x) = T(y) \Rightarrow x = y$).

<u>Theorem</u>: The following conditions ~~o~~ on a linear transformation $T: L \to L'$ are logically equivalent (i.e., any one of these conditions implies all the others).

(1) T is one-to-one.
(2) The null space of T is (0).
(3) The nullity of T is zero.
(4) If $a_1, \ldots, a_\ell$ is linearly independent in L, then $T(a_1), \ldots, T(a_\ell)$ is linearly independent in L'.
(5) If $T(x) = 0$, then $x = 0$.

<u>Proof by showing circular</u> implication $[i \Rightarrow i-1$ & $(1) \Rightarrow (5)]$

<u>(5) ⇒ (4)</u>: Assume $a_1, \ldots, a_\ell$ is linearly independent; then it must be proven, using (5), that $T(a_1), \ldots, T(a_\ell)$ is linearly independent.

$$\sum_{i=1}^{\ell} \lambda_i (T a_i) = 0 = T\left(\sum \lambda_i a_i\right)$$

But ~~s~~ (5) implies $\sum \lambda_i a_i = 0$. Since the a's are linearly independent, then $\lambda_1 = \ldots = \lambda_\ell = 0$. Hence, from the expression

$\sum_{i=1}^{\ell} \lambda_i (T a_i) = 0$, the requirement that $\lambda_1 = \ldots = \lambda_\ell = 0$ means $T(a_1), \ldots, T(a_\ell)$ is linearly independent.

(continued on page after next)

LOGIC

(1) P, Q, R mathematical propositions
(2) A proposition is either true or false.
(3) Associated with every proposition P is its negative not P.
e.g., P: $x < 7$; not P: $x \geq 7$. etc.

(4) If P and Q are related in the following way, we say P implies Q:
If P is true, then Q is also true.

(5) All the following mean (exactly) the same thing:
(a) P implies Q ; notation: $P \Rightarrow Q$ or $P \rightarrow Q$.
(b) If P, then Q.
(c) P is a sufficient condition for Q
(d) Q is a necessary condition for P.
(e) P only if Q
(f) Q, if P

~~not Q ⇒ not Q~~
not Q ⇒ not P
P ⇒ Q
are "contrapositives."

Two main facts about implication:

(1) Transitivity of implication: If $P \Rightarrow Q$ and $Q \Rightarrow R$, then $P \Rightarrow R$.

(2) $P \Rightarrow Q$ means exactly the same as not P ⇐ not Q.

"Proof by contradiction": This term is used because it follows from the fact that to prove a theorem of the form $H \Rightarrow C$, it is sufficient that not C ⇒ not H or even not C ⇒ not H^* where H^* is a consequence of H.

$H \Rightarrow H^*$ and not C ⇒ not H^* is proven

then it follows that $H \Rightarrow C$

All the following mean the same thing:

(1) $P \Rightarrow Q$ and $Q \Rightarrow P$
(2) $P \Leftrightarrow Q$
(3) P is a necessary and sufficient condition for Q
(4) P if and only if Q
(5) P and Q are logically equivalent.

Continuation of: For $T: L \to L'$, the following are logically equivalent.

(1) one-to-one
(2) Null space $= (0) = N$
(3) $n = \dim N = 0$
(4) T preserves linear independence
(5) $T(x) = 0 \Rightarrow x = 0$

Have shown $(5) \Rightarrow (4)$

$(4) \Rightarrow (3)$

$\boxed{(4) \Rightarrow (3)}$ Assume (3) to be false. Then $\dim N = n \neq 0$ so N contains a non-empty linearly independent set of vectors. Then by (4), T applied to this set would give a non-empty linearly independent set in $T(N) = (0)$. But this is impossible. Hence, a contradiction and (3) must be true.

$\boxed{(3) \Rightarrow (2)}$ Since the nullity is the dimension of the nullspace, this follows from the general fact that a vector space of dimension zero consists of the zero vector only.

$\boxed{(2) \Rightarrow (1)}$ Suppose $T(x) = T(y)$.

Then $T(x) - T(y) = 0 = T(x-y) = 0$

i.e. $(x-y)$ is in the null space of T; but (2) says $N = (0)$
Hence, $x - y = 0$ and $\therefore$ $x = y$.

$\boxed{(1) \Rightarrow (5)}$ If T is one-to-one, then since $T(0) = 0$, no other vector x has $T(x) = 0$; i.e., if $T(x) = 0 \Rightarrow x = 0$.

If $T: L \to L'$ is a linear transformation, the following conditions are logically equivalent:

(1) T onto
(2) range of $T = L'$
(3) $r = p'$ [r = rank of T ; p' = dim L'] [$r \equiv \dim R$; R = range of T]
(4) L' is spanned by a set of vectors $T(a_1), \dots, T(a_n)$

Definition: If $T: L \to L'$ is a linear transformation which is both one-to-one and onto, then T is called an isomorphism from L to L' [or between L & L']

Example: The 3 types of vectors previously considered (i.e., family, arrow, and point definitions).

Theorem: If $T: L \to L'$ is a linear transformation, then the following are logically equivalent:

(1) T is an isomorphism
(2) nullity $= 0$ and rank $= \dim L'$ (one-to-one & onto)
(3) null space $= (0)$ and range $= L'$
(4) T of a basis is a basis; i.e.; if $a_1, \dots, a_p$ is a basis for L, then $T(a_1), \dots, T(a_p)$ is a basis for L'.

Corollary: If $T: L \to L'$ is a linear transformation and $\dim L = \dim L'$ (both finite), then T is an isomorphism $\Leftrightarrow$ T onto $\Leftrightarrow$ T one-to-one.

Proof:

$$T \text{ onto} \underset{\text{def.}}{\Leftrightarrow} r = p' \underset{(p = p' \text{ by hypoth.})}{\Leftrightarrow} r = p \underset{}{\Leftrightarrow} \underset{r+n=p}{n = 0} \overset{\text{def.}}{\Leftrightarrow} T \text{ one-to-one}$$

☆ <u>Theorem</u>: Two finite dimensional vector spaces with the same scalars is an isomorphism $\Leftrightarrow$ they have the same dimension.

<u>Lemma</u>: L, L' vector spaces. Let $a_1, \dots, a_p$ be a basis for L. Let $b_1, \dots, b_p$ be <u>any</u> vectors of L'. Then there is a unique linear transformation T, such that $T(a_1) = b_1, \dots, T(a_p) = b_p$.

(This says that a linear transformation is characterized by its effect on a basis.)

<u>Proof</u>: 2 parts: (1) existence, &, (2) uniqueness.

(2) If T and T' are linear transformations $(L \to L')$ carrying a_i to b_i $(1 \le i \le p)$, then claimed that $T = T'$. That is, $T(x) = T'(x)$ for all x in L. If $x \in L$, we can write $x = \sum \lambda_i a_i$ because $a_1, \dots, a_p$ span.

Thus, $$T(\sum \lambda_i a_i) = \sum \lambda_i T(a_i) \quad \text{due to linearity.}$$
$$= \sum \lambda_i b_i$$

$$T'(\sum \lambda_i a_i) = \sum \lambda_i T'(a_i) = \sum \lambda_i b_i \qquad \text{Q.E.D.}$$

(1) Define a function T that takes $L \to L'$. If $x \in L$, write $x = \sum \lambda_i a_i$. λ_i are <u>uniquely</u> determined for a given x. Then let $T(x) = \sum \lambda_i b_i$. Since λ_i's are unique, so is $T(x)$ and hence, T is a function

<u>Proof of Theorem</u>: L and L' are vectorspaces. P & P' are dim of L & L' respectively.

(1) $P = P' \Rightarrow L$ isomorphic with L'.

Pick a basis $a_1, \dots, a_p$ for L and $b_1, \dots, b_p$ for L'. By the Lemma, there exists a linear transformation $T: L \to L'$ such that $T(a_1), \dots, T(a_p) = b_1, \dots, b_p$

Since $T(a_1), \dots, T(a_p)$ span, T is onto but since L and L' have the same dimension $(P = P')$, T onto $\Rightarrow T$ is isomorphic.

(2) Isomorphism $\Rightarrow$ same dimension

If $T: L \to L'$ is an isomorphism, then T carries a basis for L to a basis for L'. Since the number of vectors in any basis of $V = \dim V$, L & L' must have the same dimension.

<u>Corollary</u>: Every finite dimensional vector space over the real numbers (scalars = reals) is isomorphic with an R_p.

<u>Proof</u>: If $\dim L = P$, then by the theorem, L is isomorphic with R_P.

<u>COMPUTATIONS IN R_p</u>

Start work on a method of computing the dimension of the subspace R_q spanned by P given vectors $a_1, \dots, a_p$ in R_q.

<u>Definition</u>: A $p \times q$ <u>matrix</u> (over the real or complex numbers) is a rectangular array of numbers, thus:

$$\begin{pmatrix} a_{11}, & \alpha_{12} & \alpha_{13} & \dots & \alpha_{1q} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \dots & \alpha_{2q} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \dots & \alpha_{3q} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \alpha_{p1} & \alpha_{p2} & a_{p3} & \dots & \alpha_{pq} \end{pmatrix}$$

The horizontal q-tuples in a matrix are called its rows (vectors in R_q).
The vertical p-tuples are called columns (vectors in R_p).
The numbers α_{ij} are called entries.

The standard notation for a matrix (see above) is: The numbers α_{ij} (where $1 \le i \le P$, $1 \le j \le q$) describe a $P \times q$ matrix with α_{ij} in the i^{th} row and the j^{th} column.

<u>Definition</u>: Elementary row operations on a matrix:

(1) Interchange any two rows.
(2) Multiply one row by a non-zero scalar
(3) Add one row to a different row.

Given p vectors $a_1, \ldots, a_p$ in R_q; they determine a $p \times q$ matrix as follows: Let the q-tuple a_i ~~(...)~~ be the i^{th} row of the matrix. So by convention about index notation, must write $a_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iq})$ for $1 \le i \le P$. Then α_{ij} is the matrix.

Theorem: If row operations are performed on a $p \times q$ matrix, the subspace of R_q spanned by the ~~[illegible]~~ rows of the matrix does not change.

Proof: For operation of interchanging rows and scalar multiplying rows, the theorem is obvious since (1) order is immaterial and (2) if $a_1, \ldots, a_p$ span, so do $\lambda a_1, \ldots, \lambda a_p$.

B) For the operation of addition of rows, let $a_1, \ldots, a_p$ be the rows and let them span R_p. Then, to prove $a_1 + a_2, a_2, \ldots, a_p$ span.

Let any vector v in R_q be $v = \sum_{i=1}^{p} \lambda_i a_i = \lambda_1 a_1 + \lambda_2 a_2 + \sum_{i=3}^{p} \lambda_i a_i$.

Then by letting $v = \mu_1(a_1 + a_2) + \mu_2 a_2 + \sum_{i=3}^{p} \mu_i a_i$ means that if $\mu_1 = \lambda_1$ and $\mu_2 + \mu_1 = \lambda_2$ and $\mu_i = \lambda_i \ (3 \le i \le P)$ then $a_1 + a_2, a_2, \ldots, a_p$ span. These conditions on μ_i are easily satisfied for given λ_1 and λ_2; hence, ~~they~~ $a_1 + a_2, a_2, \ldots, a_p$ span.

Definition: A $p \times q$ matrix is in (row) echelon form if each row begins with more zeroes than the preceding rows. (Not necessarily only one more).

Theorem: If a matrix is in echelon form, then its non-zero rows are linearly independent.

$$\left.\begin{array}{llllllll} \text{Write the first row as} & \lambda e & \lambda f & \lambda g & \ldots & \lambda h \\ \text{second} \quad " & " & 0 & a & b & \ldots & c \\ n^{th} \quad " & " & 0 & 0 & 0 & \ldots & d \end{array}\right\} \text{ with } \lambda \ne 0 \ne e \text{ in particular}$$

The first row cannot be written as $\sum \mu_i \, \text{row}_i$ unless $\lambda = 0$ which is contrary to assumption since all other rows have zeroes in the first column. Similarly, the second row cannot be thus expressed unless $\lambda \mu_1 e = 0 \Rightarrow \mu = 0$. Similarly for the 3^{rd} row $\rightarrow$ ~~$\lambda \mu_1 f + \mu$~~ Since μ_1 must be zero, $\mu_1 \lambda f + \mu_2 a = 0$ requires $\mu_2 = 0$. Etcetera ad nauseum et infinitum.

Method of computing the dimension of the subspace S of R_q spanned by the given vectors $a_1, \ldots, a_p$:

(1) Assemble ~~$a_1, \ldots, a_p$~~ these ~~vectors~~ n-tuples into a $p \times q$ matrix
(2) Using the elementary row operations, reduce the above (1) matrix into echelon form.
(3) Count the non-zero ~~vectors~~ rows in the echeloned matrix

This number is the required dimension.

To ~~so~~ arrange a matrix into echelon form: Take the first non-zero column and by interchanging rows, move ~~the~~ a row with ~~the~~ a non-zero entry in that column to the first row. Then, using scalar multiplication and addition of rows, eliminate all non-zero entries in this column below the first row. Next, go to the next column which has a non-zero entry and by using the row operations on any of the rows below the preceding one, place the row with the non-zero entry in the first row considered (i.e., immediately below the preceding row and reduce all entries in this column and below the considered row to zero. Et cetera, mutatis mutandis.

EXAMPLE

$$\begin{pmatrix} 1 & -1 & 0 \\ 2 & -1 & 1 \\ -1 & 3 & -2 \\ -1 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 2 & -2 \\ 0 & -1 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & -4 \\ 0 & 0 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & -4 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow \dim S = 3$$

i.e. $S = R_n$

Consequences of these calculations:

(1) Can compute whether vectors $a_1, \ldots, a_p$ which span S in R_q are linearly independent since they are linearly independent if & only if $\dim S = p$.

(2) New proof that $\dim R_n = n$ since can show that more than n vectors in R_n must be linearly dependent.

Now, we begin to compute with linear transformations. Before doing so, we note that connections exist between:

(1) a linear transformation $A : R_u \to R_n$
(2) $n \times u$ matrix (α_{ij}) where $1 \le i \le n$ and $1 \le j \le u$.
(3) a certain set of n simultaneous linear equations ~~in the~~ in u unknowns $\xi_1, \dots, \xi_u$.

<u>Definition</u>: Given a linear transformation $A : R_u \to R_n$, the <u>matrix of A</u> is the $n \times u$ matrix whose <u>columns</u> (in natural order) are the n-tuples $A(e_1), \dots, A(e_u)$ where $e_i = (0, \dots, 1, \dots, 0)$ where the 1 is the i^{th} term.

<u>A note on notation</u>: As a result of the convention that (α_{ij}) has i as the row index and j as the column index, we must write $A(e_i) = (\alpha_{1i}, \alpha_{2i}, \dots, \alpha_{ni})$; i.e., this is the i^{th} column of the matrix of A.
Since $A(e_i)$ is to be the i^{th} column of the matrix of A, the index i must appear ~~second~~ in the second (or column) position. So actually, the matrix of $A = (\alpha_{ki})$.

$$\left[\begin{array}{l} M(n \times u) \equiv \text{the set of all } n \times u \text{ matrices} \\ \mathcal{L}(R_u, R_n) \equiv \text{the set of all linear transformations carrying } R_u \text{ to } R_n \end{array}\right.$$

Computation with linear transformations: (Matrix notation previously defined). Aim to show that (in a sense to be made precise) matrices and linear transformations are essentially equal.

<u>Def</u>: Let $T : R_u \to R_n$ be a linear transformation from R_u to R_n. Let $e_1, \dots, e_u$ be the unit vectors in R_u. Then the <u>matrix of T</u> (A_T) is the $n \times u$ matrix whose j^{th} column is the n-tuple $T(e_j)$

<u>Notation</u>: In the general case of $T : R_u \to R_n$, the notation used for $T(e_j)$ is completely determined by the definition of A_T and the matrix notation (α_{ij}) where i = column index ; j = row index

Thus, if A_T is to be (α_{ij}), we must write $T(e_j) = (\alpha_{1j}, \alpha_{2j}, \dots, \alpha_{nj})$

<u>Definition</u>: Let $A = (\alpha_{ij})$ be a $n \times u$ matrix
Let $T_A : R_u \to R_n$ be the function such that

$$T_A(\xi_1, \dots, \xi_u) = \left(\sum_{j=1}^{u} \alpha_{1j}\xi_j , \dots , \sum_{j=1}^{u} \alpha_{nj}\xi_j \right)$$

Then T_A is linear and is called the <u>transformation of A</u>

T_A can be described as: left matrix-multiplication of column vectors by the matrix A:

$$\begin{pmatrix} \alpha_{11}, & \alpha_{12}, & \dots, & \alpha_{1u} \\ \alpha_{21}, & \alpha_{22}, & \dots, & \alpha_{2u} \\ \cdot & \cdot & \cdot & \cdot \\ \alpha_{n1}, & \alpha_{n2}, & \dots, & \alpha_{nu} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_u \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}$$

where $n > u$

$\hookrightarrow T(\xi_1, \dots, \xi_u)$

where T is the linear transformation associated with the $n \times u$ matrix

i^{th} component of T_A is: $\alpha_{i1}\xi_1 + \alpha_{i2}\xi_2 + \dots + \alpha_{iu}\xi_u$

$$\begin{tikzcd} \boxed{\begin{array}{l}\text{Set of all linear transformations} \\ \text{from } R_u \to R_n \\ \mathcal{L}(R_u, R_n)\end{array}} \arrow[r, shift left, "T_A \quad A"] & \boxed{\begin{array}{l}\text{Set of all } n \times u \\ \text{matrices} \\ M(n \times u)\end{array}} \arrow[l, shift left, "T \quad A_T"] \end{tikzcd}$$

The two functions (correspondences) carrying $T \to T_A$ and $A \to A_T$ both define the same one-to-one correspondence between $\mathcal{L}(R_u, R_n)$ and $M(n \times u)$. [This is expressed in the theorem in homework problem #7].

So, we now know that the linear transformations from R_u to R_n are all given by the $n \times u$ matrices acting by left matrix-multiplication of column vectors.

Def: If A is an $n \times u$ matrix, define

(1) Row rank = dimension of subspace of R_u spanned by the n rows of A.
(2) Column rank = " " " " R_n " " " u columns of A.
(3) Rank = the rank of A as a linear transformation.

Note: Row rank can be computed by row echeloning.
Column " " " " " column " .

Rank = column rank because $A(e_j)$ is just the j^{th} column of A.
(Here, we use A to denote both the matrix and the linear transformation associated with it.)
The $A(e_j)$'s span the range.

To be computed for a linear transformation $A: R_u \to R_n$

(1) rank
(2) nullity
(3) Conditions that a given $b \in R_n$ is in the range.
(4) Basis for null space
(5) Basis for range

Methods of computation

(1) rank = column rank; so reduce matrix to echelon form by column operations and number of non-zero columns = rank.
(2) nullity = u - rank
(3) Range is spanned by $A(e_1), \ldots, A(e_u)$ and $A(e_j) = j^{th}$ column of A; so $b \in$ range $= [A(e_1), \ldots, A(e_u)] \Leftrightarrow$ rank $= \dim [A(e_1), \ldots, A(e_u)]$ $= [A(e_1), \ldots, A(e_u), b]$

(4) later
(5) Column echelon; ~~then~~ non-zero columns (in each matrix) are a basis for the range.

Linear Equations (Application of vector space theory)

Given the numbers α_{ij} $(1 \le i \le n ; 1 \le j \le u)$ and $\beta_1, \ldots, \beta_n$, consider

$$\begin{aligned} \alpha_{11}\xi_1 + \alpha_{12}\xi_2 + \cdots + \alpha_{1u}\xi_u &= \beta_1 \\ \alpha_{21}\xi_1 + \alpha_{22}\xi_2 + \cdots + \alpha_{2u}\xi_u &= \beta_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \alpha_{n1}\xi_1 + \alpha_{n2}\xi_2 + \cdots + \alpha_{nu}\xi_u &= \beta_n \end{aligned}$$

called n simultaneous linear equations in u unknowns. If $\xi_1, \ldots, \xi_u$ are numbers which satisfy these equations, these numbers are called a solution.

If $\beta_1 = \ldots = \beta_n = 0$, equations are called homogeneous; otherwise, non-homog.

Translate all this into vector space terminology & notation. The α_{ij}'s form an $n \times u$ matrix, and hence, a linear transformation from R_u to R_n. The numbers $\beta_1, \ldots, \beta_n$ form a vector $b = (\beta_1, \ldots, \beta_n) \in R_n$.

The above equations can be written as $A(x) = b$. A solution $x = (\xi_1, \ldots, \xi_n)$ is a vector in R_u for which this equation holds; i.e., the linear transformation A carries the vector x to the given vector b.

Example:

$u = 4$
$n = 3$

$$\begin{aligned} 2\xi_1 - \xi_2 + 4\xi_4 &= 1 \\ \xi_1 + \xi_2 + \xi_3 &= -1 \\ \xi_1 - 8\xi_2 + 6\xi_4 &= 0 \end{aligned}$$

$$\rightarrow \begin{pmatrix} 2 & -1 & 0 & 4 \\ 1 & 1 & 1 & 0 \\ 1 & -8 & 0 & 6 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

Consider a set of n equations (homogeneous linear) in u unknowns, $A(x) = Ax = 0$. A is an $n \times u$ matrix of coefficients and is considered as a linear transformation $A: R_u \to R_n$. Hence, two fundamental & easy facts:

(1) The set of solutions to $Ax = 0$ forms a subspace of R_u; eg., the null space of R_u. So dim of solution subspace to $Ax = 0$ (call it d) = nullity of A. We can compute nullity (rank = number of non-zero vectors in column echelon form). Then $u - r$ = nullity. Note if $d > 0$ (rank of $A < u$), we have non-trivial solutions + trivial 0 to $Ax = 0$. In particular, since rank $A \le n$, if $n < u$, we have non-trivial solutions.

(2) Elementary row-operations are defined for a set of equations in natural way-

1.) Rearrange equations
2.) Multiply an equation by $\lambda \neq 0$
3.) Add one equation to another (different) equation

The effect of these operations is the same as performing the corresponding row operations on matrix A, and it does not change the solution subspace.

<u>Proof</u>: Write out $Ax = 0$. It is clear that (1) and (2) do not change the solution subspace. Add first equation to second (now perfectly general case by (1)) and show that $x = \xi_1, \dots, \xi_u$ is a solution to the original equation $\Leftrightarrow$ it is a solution to added equation.

<u>Def</u>: <u>Super-echelon form</u>: By elementary row operations we can get a given matrix into echelon form - and by multiplying suitable rows by suitable scalars, we can make each lead coefficient equal to $+1$.

Then take the last non zero row, and use it to get zero's in the column containing its first non-zero term (which is now 1) Continue similarly, ...

$$\begin{pmatrix} 1 & \alpha & \alpha & 0 & 0 & \alpha & \alpha & 0 & \alpha \\ 0 & 0 & 0 & 1 & 0 & \alpha & \alpha & 0 & \alpha \\ 0 & 0 & 0 & 0 & 1 & \alpha & \alpha & 0 & \alpha \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \alpha \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Matrix in super-echelon form

$n = 5$, $u = 9$

$r = 4$,

$t_1 = 1, t_2 = 4, t_3 = 5, \ t_4 = 8$

<u>Theorem</u>: Given an $m \times n$ matrix, by elementary row operations, we can obtain an $m \times n$ matrix in super-echelon form, i.e., with following properties:

(1) The first r rows of (α_{ij}) are non-zero, the rest are zero.

(2) There exist integers $1 \leq t_1 < t_2 < \dots < t_r \leq n$ such that:

(a) $\alpha_{i t_i} = 1$ for $i = 1, 2, \dots, r$

(b) $\alpha_{i t_i}$ is the first non-zero entry in the i^{th} row

(c) $\alpha_{i t_i}$ is the only non-zero entry in the t_i^{th} column.

Method of solving $A(x) = 0$. A is $m \times n$

1. ~~Let~~ Put A into super echelon form (call it A') by row operations with notation as in previous theorem.

2. Solve $A'(x) = 0$ and get solution of $Ax = 0$. Write $A'(x)$ as set of equations. Each ξ_{t_i} $(i = 1, \dots, r)$ occurs in one and only one equation with coefficient one. Pick any numbers for ξ_j, where $j \neq t_1 \neq t_2 \neq \dots \neq t_r$. We can trivially solve $A'(x) = 0$ for ξ_i $(i = t_1, \dots, t_r)$

<u>METHOD OF SOLVING LINEAR HOMOGENEOUS EQUATIONS</u>

Given $A(x) = 0$

(I) Super echelon A, getting $A'(x) = 0$; where A' has r non-zero rows with integers $t_1, \dots, t_r$.

(II) Solve for $\xi_{t_1}, \xi_{t_2}, \dots, \xi_{t_r}$ (each of these is only one equation) after picking any values at all for the other variables.

<u>EXAMPLE</u>:

$$\begin{aligned} \xi_1 + \xi_2 + \xi_3 + \xi_4 &= 0 \\ 2\xi_1 + 3\xi_2 - \xi_3 + \xi_4 &= 0 \\ 3\xi_1 + 4\xi_2 \qquad + 2\xi_4 &= 0 \end{aligned} \Leftrightarrow \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 3 & -1 & 1 \\ 3 & 4 & 0 & 2 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} 1 & 0 & 4 & 2 \\ 0 & 1 & -3 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{pmatrix} \Leftrightarrow \begin{aligned} \xi_1 \qquad + 4\xi_3 + 2\xi_4 &= 0 \\ \xi_2 \, \cancel{+} - 3\xi_3 - \xi_4 &= 0 \end{aligned} \Rightarrow \begin{aligned} \xi_1 &= -4\xi_3 - 2\xi_4 \\ \xi_2 &= 3\xi_3 + \xi_4 \end{aligned}$$

To get a basis for the solution subspace: ~~Let~~

$$\left. \begin{aligned} &\text{Let } \xi_3 = 1, \xi_4 = 0 \Rightarrow \xi_1 = -4, \xi_2 = +3 \\ &\text{Let } \xi_3 = 0, \xi_4 = 1 \Rightarrow \xi_1 = -2, \xi_2 = 1 \end{aligned} \right\} \text{So } (-4, 3, 1, 0) + (-2, 1, 0, 1) \text{ are a solution to } A(x) = 0$$

(III) <u>To get a basis for solution subspace:</u>

In what follows, $1 \le k \le u$; $k \ne t_1 \ne t_2 \ne \cdots \ne t_r$; k is an integer
k will tell which undetermined unknowns we have.
Note there are $u-r$ of these, r = row rank of matrix A.

For each k, define vector $X_k \in R_u$ as follows:

Let

$$X_k = (\xi_{k_1}, \dots, \xi_{k_u}) \text{ where } \xi_{k_j} = \begin{cases} 1 \text{ if } j=k \\ 0 \text{ if } j \ne k ;\ j \ne t_1, \dots, t_r \end{cases} \text{ undetermined unknowns}$$
$$\qquad \text{Solve equations if } j = t_1, \dots, t_r \ \text{ determined unknowns.}$$

These X_k's are solutions of $A'(x) = 0$ (and hence, of $A(x) = 0$)

* {<u>They are a basis of the subspace</u>
{Proof:

Suppose $\sum_k \lambda_k X_k = 0$. Then, in particular, the j^{th} component of this vector is zero. $\rightarrow j \ne t_1, \dots, t_r$

$$0 = \Big(\sum_k \lambda_k X_k\Big)_j = \sum_k \lambda_k \xi_j$$

Now, since k and j aren't t_i's, for fixed j, the sum above reduces to one term only; namely, $\xi_{jj} = 1$
So equation above is: <u>$0 = \lambda_j$</u> ; $1 \le j \le u$, $j \ne t_1 \ne \cdots \ne t_r$
Hence, <u>independence</u>

Let $c = (\gamma_1, \dots, \gamma_u)$ be a solution. Claim that $c = \sum_k \gamma_k X_k$
i.e., show $c - \sum_k \gamma_k X_k = 0$. Now notice this vector is a solution.

<u>Lemma</u>: If $s = (\sigma_1, \dots, \sigma_u)$ is a solution and $\sigma_j = 0$ for $j \ne t_i$ then whole vector s is zero [$s = 0$]

<u>Proof</u>: Because $\sigma_{t_1}, \dots, \sigma_{t_r}$ are linear combinations of the σ_j's, where $j \ne t_1 \ne t_2 \ne \cdots \ne t_r$.

Hence, to prove it is zero, need only prove undetermined components are zero.

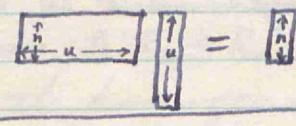
i.e., If $j \ne t_1 \ne \cdots \ne t_r$, consider $\Big(c - \sum_k \gamma_k X_k\Big)_j = \gamma_j - \sum_k \gamma_k \xi_{kj}$

$k \ne t_1 \ne \cdots \ne t_r$
$j \ne t_1 \ne \cdots \ne t_r$

$= \gamma_j - \gamma_j = 0$

when $k=j$

REVIEW: Studying linear transformations $A: R_u \to R_n$;
Basic fact is that all linear transformations are gotten by matrix multiplication ($n \times u$ matrices) of column vectors



Row rank of matrix A = dim of subspace of R_u spanned by rows of A.
Column rank " " " " " " " R_n " "columns" ".

Want to prove row rank = column rank.
Method ~~of solution~~ will be to prove (column rank) $\overset{\text{def.}}{=}$ (rank of A as a linear transformation) = (row rank). Specifically, the method of solutions:

(1) By row operations, reduce A to super echelon form $A'(x)=0$ with characteristic integers $t_1 < \dots < t_r$ where r = number of non-zero rows of A' [$Ax=0$ and $A'x=0$ have same solutions].

(2) Trivially solve $A'x=0$ for $\xi_{t_1}, \dots, \xi_{t_r}$. Any choice of other ξ's gives a solution to $Ax=0$. In fact, if the undetermined ξ's, i.e., ξ_i ($i \neq t_1 \neq \dots \neq t_r$), are chosen in the usual way (all zeroes except one number, which is one [similar to unit vectors]), then the result is $u-r$ vectors forming a basis for the solution subspace of $Ax=0$.

Theorem: If A is an $n \times u$ matrix, then (column rank of A)
= (rank of A considered as a linear transformation from R_u to R_n)
= (row rank of A).

(Proof over)

Proof: First equality follows immediately from definition of rank & column rank [cf. p. 24].

Let $t_1, \ldots, t_r$ and r have usual significance in a super-echeloning of $Ax = 0$ to $A'x = 0$

(1) r is the row rank of A because it has been proven that any echeloning (in particular, super-echeloning) of A leaves a number of non-zero rows equal to the row rank.

(2) By the last feature of the method of solving $Ax = 0$ (i.e., the statement that a basis for the solution subspace consists of $u-r$ vectors), we know that the dimension of the solution subspace of $Ax = 0$ is $u-r$; but the null space of A is the solution subspace of $Ax=0$; so $\underline{u-r = \text{nullity}}$. But rank of A + nullity of A = $u = \dim R_u$
Hence, r = rank of A.

(1) and (2) combined prove row rank = rank & hence,
row rank = column rank = rank.

Consequences of the theorem:

To find the rank of a matrix, either row or column echeloning can be used. (count non-zero rows or columns).
So, given $A(x) = 0$, the dimension of the solution subspace can be found as $u-r$, r = rank of A.
Non-trivial solutions exist if nullity > 0 or $r < u$.

Non-homogeneous linear equations

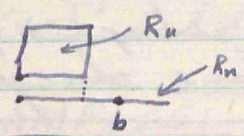
Problem here is (in matrix-linear transformation terms):

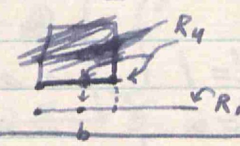
Given $A: R_u \to R_n$ and $b \in R_n$, find all vectors $x \ni Ax = b$. In the homogeneous case, either one solution [necessarily the trivial solution $(0,0,\ldots,0)$] or an infinite number of solutions are found. In the non-homogeneous case, there are three possibilities:

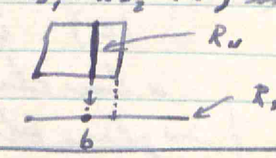
(1) no solution ; b not in range
(2) one solution ; b in range, A one-to-one
(3) infinitely many solutions ; b in range, A not one-to-one.

EXAMPLE: $n = u = 2$

$\left.\begin{array}{l}\xi_1 + \xi_2 = 6 \\ \xi_1 + \xi_2 = 2\end{array}\right\}$ no solution ; $\left.\begin{array}{l}\xi_1 + \xi_2 = 6 \\ \xi_1 - \xi_2 = 0\end{array}\right\}$ one solution ; $\left.\begin{array}{l}\xi_1 + \xi_2 = 6 \\ 2\xi_1 + 2\xi_2 = 12\end{array}\right\}$ infinitely many solutions







Theorem: Given a linear transformation $A: R_u \to R_n$ and $b \in R_n$, $(b \neq 0)$, consider $Ax = b$. Let y_0 be a particular solution to this non-homogeneous equation. Then a vector y in R_u is a solution to this non-homogeneous equation if and only if y is of the form $y_0 + h$ where h is a solution to the corresponding homogeneous equation $A(x) = 0$.

Proof: Assume y is a solution to $Ax = b$. Then $Ay = b$ and $Ay_0 = b$; hence $A(y - y_0) = A(y) - A(y_0) = b - b = 0$. Thus, $y - y_0$ is a solution to corresponding homogeneous equation $Ax = 0$. Let $h = y - y_0$; then $y = y_0 + h$.

Assume $y = y_0 + h$ where $A(y_0) = b$ and $A(h) = 0$.

$A(y_0 + h) = Ay_0 + Ah = b + 0 = b$

Algebraic significance of theorems:

If ~~only~~ one solution to the non homogeneous ~~solution~~ equation can be found, all the rest can be found by considering only the presumably easier homogeneous equation.

If $h_1, \dots, h_k$ is a basis for the solution subspace of $Ax = 0$, then every solution of $Ax = b$ is of the form $y = y_0 + \lambda_1 h_1 + \dots + \lambda_k h_k$ where $\lambda_1, \dots, \lambda_k$ are completely arbitrary. This is sometimes called the ~~non~~ general solution of the non-homogeneous equation. $\lambda_1 h_1 + \dots + \lambda_k h_k$ is the general solution of the corresponding homogeneous equation.

Geometric significance:

Solutions to $Ax = 0$ form the null space of A, i.e., a subspace of R_n. For example, in R_3 a subspace of dimension d is:

$$\left.\begin{array}{l} (0) \text{ if } d = 0 \\ \text{a line through } 0 \text{ if } d = 1 \\ \text{a plane through } 0 \text{ if } d = 2 \\ R_3 \text{ if } d = 3 \end{array}\right\}$$

In general, a subspace of R_n is a hyper-plane through 0.

Theorem says that the solution of $Ax = b$, i.e., the set of all vectors in R_n carried to b by A, form a hyper-plane through y_0 (assuming $Ay_0 = b$) [because the geometric effect of adding y_0 to the solution subspace $Ax = 0$ is to displace this hyperplane parallel to itself till it contains y_0].

<u>Theorem</u>: Given the equation $Ax = b$, the elementary row operations performed on the equations (including the non-homogeneous terms β_i, $1 \le i \le n$) do not change the solutions.

<u>Method of solution (if possible) of $Ax = b$</u>

(1) By elementary row operations on equations (including β's), get A into super-echelon form A'. So, $A'x = b'$ where A' has $t_1, \dots, t_r$ and r with usual significance.

(2) <u>Case (I)</u>: If b' has a component $\beta_i' \neq 0$ with $i > r$, then equations $A'x = b'$ cannot be solved because have $0 = \beta_i' \neq 0$. Hence, $Ax = b$ cannot be solved.

<u>Case (II)</u>: Otherwise, i.e., if $\beta_i = 0$ for $i > r$, then proceed as before, i.e., take any values [unit vector analogy] for ξ_i ($i \neq t_1, \dots, t_r$) and solve $A'x = b'$ for $\xi_{t_1}, \dots, \xi_{t_r}$ to get a solution vector $(\xi_1, \dots, \xi_n)$ to $A'x = b'$ and hence, to $Ax = b$.

<u>EXAMPLES</u>: ①

$$\begin{aligned} \xi_1 - 2\xi_2 + \xi_3 &= 2 \\ 2\xi_1 \qquad + \xi_3 &= 0 \\ \xi_1 - 2\xi_2 - \xi_3 &= -6 \end{aligned} \Rightarrow \begin{pmatrix} 1 & -2 & 1 & 2 \\ 2 & 0 & 1 & 0 \\ 1 & -2 & -1 & -6 \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} 1 & -2 & 1 & 2 \\ 0 & 4 & -1 & -4 \\ 0 & 0 & -2 & -8 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & -\frac{1}{4} & -1 \\ 0 & 0 & 1 & 4 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 4 \end{pmatrix}$$

$$\boxed{\xi_1 = -2 \; ; \; \xi_2 = 0, \; \xi_3 = 4}$$

②

$$\begin{aligned} 2\xi_1 + \xi_2 + 3\xi_3 &= 4 \\ \xi_1 + \xi_2 + \xi_3 &= 2 \\ \xi_1 + 2\xi_2 \qquad &= 1 \end{aligned} \Rightarrow \begin{pmatrix} 2 & 1 & 3 & 4 \\ 1 & 1 & 1 & 2 \\ 1 & 2 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & \frac{1}{2} & \frac{3}{2} & 2 \\ 0 & +1 & 1 & 2 \\ 0 & 1 & -1 & -1 \end{pmatrix}$$

$$\Rightarrow \cancel{\begin{pmatrix} 1 & 0 & 2 & \frac{5}{2} \\ 0 & 1 & +1 & +2 \\ 0 & -1 & -1 & -1 \end{pmatrix}} \Rightarrow \begin{pmatrix} 1 & \frac{1}{2} & \frac{3}{2} & 2 \\ 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 1 & -1 & -1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 2 & 2 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & -1 \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} 1 & 0 & 2 & 2 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$ $\therefore$ no solution since last equation becomes $0 = -1$

<u>Example</u> : $\left.\begin{array}{l}\xi_1 + 3\xi_2 - \xi_3 = 1 \\ 2\xi_1 + 2\xi_2 - 2\xi_3 = 1 \\ 3\xi_1 + \xi_2 - 3\xi_3 = 1\end{array}\right\}$ $\left.\begin{array}{l}\xi_1 - \xi_3 = 1/4 \\ \xi_2 = 1/4\end{array}\right\}$ $\begin{array}{l}\xi_1 = \xi_3 + 1/4 \\ \xi_2 = \frac{1}{4}\end{array}$

Solution is of the form $(t+1/4, 1/4, t)$ or $\underbrace{(1/4, 1/4, 0)}_{\text{particular sol. to } Ax=b=(1,1,1)} + \underbrace{(t, 0, t)}_{\text{general solution to } Ax=0=(0,0,0)}$

Given a non-homogeneous equation $Ax=b$, form a matrix by adding $b=(\beta_1, \dots, \beta_n)$ to the matrix of coefficients as a $(n+1)^{st}$ column. This $n \times (n+1)$ matrix is called the <u>augmented</u> <u>matrix</u> of the equations $Ax=b$; denoted by A^*.

$$A = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \cdot & \cdot & \cdot \\ \alpha_{n1} & \cdots & \alpha_{nn} \end{pmatrix} \longrightarrow A^* = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} & \beta_1 \\ \cdot & \cdots & \cdot & \cdot \\ \alpha_{n1} & \cdots & \alpha_{nn} & \beta_n \end{pmatrix}$$

<u>Theorem</u> : $Ax=b$ has solutions $\Leftrightarrow$ rank A = rank A^*

<u>**Proof**</u>:

$Ax=b$ has solutions $\overset{\Longleftrightarrow}{\text{if and only if}}$ $b \in$ range of A $\Leftrightarrow$ $b \in [Ae_1, \dots, Ae_n]$. But $Ae_j = j^{th}$ column of A so can calculate $\dim[Ae_1, \dots, Ae_n]$ as column rank of A. Similarly, $\dim[Ae_1, \dots, Ae_n, b]$ = column rank of augmented matrix A^*. But column rank = row rank = rank. So conditions above are all true if and only if rank A = rank A^*

$$\text{If rank } A^* > \text{rank } A \Rightarrow \begin{pmatrix} \cdots\cdots & \beta_1 \\ \cdots\cdots & \beta_2 \\ 0\,0\,0\,0\,0 & \beta_n \end{pmatrix} \Rightarrow 0 = \beta_n \neq 0.$$

Hence, rank $A^* \leq$ rank A. But rank $A^* \geq$ rank A intuitively. Hence, rank A = rank A^*.

CORRECTION : Method of solving $Ax = b$.

1. Using row operations, super echelon augmented matrix A^*

2. Case I : If there is a $\beta_i' \neq 0$ & $i > r$, then no solutions exist.

Case II : If for $i > r$, $\beta_i = 0$, can solve $A'x = b'$: Pick undetermined ξ_i's arbitrarily (perhaps all $= 0$) and solve equation for a particular solution: $y_0 = (\xi_1, \ldots, \xi_4)$

Then solve corresponding homogeneous equation $A'x = 0$. Note that $A'x = 0$ is merely super-echeloned $Ax = 0$. Solve these equations as before (trivially) and get a basis x_i ($i \neq t_1, \ldots, t_r$) for solution subspace $Ax = 0$

Then $y = y_0 + \sum_{i = t_1, \ldots, t_r} \lambda_i x_i$ is general solution, i.e., for appropriate λ_i, all solutions are found.

EXAMPLE:

$$A^* = \begin{pmatrix} 1 & -2 & 0 & 3 & 3 \\ 2 & -4 & 1 & 3 & 7 \\ 1 & -2 & 1 & 0 & 4 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & -2 & 0 & 3 & 3 \\ 0 & 0 & 1 & -3 & 1 \\ 0 & 0 & 1 & -3 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & -2 & 0 & 3 & 3 \\ 0 & 0 & 1 & -3 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \left.\begin{matrix} r = 2 \\ t_1 = 1; t_2 = 3 \end{matrix}\right\} \text{equation have solution.}$$

ξ_1 & ξ_3 are determined; ξ_2 & ξ_4 are undetermined.

$$\left.\begin{aligned} \xi_1 - 2\xi_2 + 3\xi_4 &= 3 \\ \xi_3 - 3\xi_4 &= 1 \end{aligned}\right\} \quad \text{To get } y_0, \text{ let } \xi_2 = \xi_4 = 0 \Rightarrow \xi_1 = 3,\ \xi_3 = 1$$

$$\Rightarrow \quad y_0 = (3, 0, 1, 0)$$

Checking with original equations → $3 = 3$; $6 + 1 = 7$; $3 + 1 = 4$ ⟵

Now, find basis for solution to corresponding $A'x = 0$

$$\left.\begin{aligned} \xi_1 - 2\xi_2 + 3\xi_4 &= 0 \\ \xi_3 - 3\xi_4 &= 1 \end{aligned}\right\} \quad \text{Basis } x_i\ (i \neq t_1, t_2) = x_2,\ x_4$$

$\xi_2 = 1$, $\xi_4 = 0 \rightarrow (x_2) \Rightarrow \xi_1 - 2 = 0$, $\xi_3 = 0 \Rightarrow x_2 = (2, 1, 0, 0)$

$\xi_2 = 0$, $\xi_4 = 1 \rightarrow (x_4) \Rightarrow \xi_1 + 3 = 0$, $\xi_3 - 3 = 0 \Rightarrow x_4 = (-3, 0, 3, 1)$

Check to see if x_2 & x_4 __are__ solutions to $A\vec{x} = 0$

$2-2=0$
$4-4=0$ ✓
$2-2=0$

$-3+3=0$
~~$\ldots$~~ $= -6+3+3=0$ ✓
$-3+3=0$

Therefore, every solution to $Ax = b$ is of form

$$y = (3,0,1,0) + \lambda_2(2,1,0,0) + \lambda_4(-3,0,3,1)$$

~~M.M. on p. 36~~ on p. 42

Composition of linear transformations & matrix multiplication

Def: If $S: L \to L'$ and $T: L' \to L''$ are linear transformations, then let $TS: L \to L''$ be the function $\ni (TS)(x) = T(S(x))$; TS is a linear transformation called the composition of S & T.

NOTE: L' is called the codomain of $T \circ S$.

$$\underset{\cdot x}{L} \xrightarrow{S} \underset{\cdot Sx}{L'} \xrightarrow{T} \underset{\cdot TSx}{L''}$$

Def: Let $T: L \to L'$ be an isomorphism. If $y \in L'$, then since T is onto, $\exists\, x \in L \ni Tx = y$. Since T is one-to-one, there is only one such vector. Then let T^{-1} be the function that sends y to this x. $T^{-1}: L' \to L$ is linear, and, in fact, an isomorphism called the inverse of T.

Elementary properties:

(1) If T is an isomorphism, then $(T^{-1})^{-1} = T$

(2) If $S: L \to L'$ & $T: L' \to L$ & if:

(a) TS = identity transformation of L
[$I: L \to L \ni I(x) = x$ is the identity transformation of L]

(b) ST is the identity transformation of L',

then S & T are inverse isomorphisms.

Matrices & Linear Transformations $L \to L'$

We have a one-to-one correspondence between $\mathcal{L}(R_u, R_n) \leftrightarrow M(n \times u)$. We want a similar one: $\mathcal{L}(V, W) \leftrightarrow M(n \times u)$, where V & W are arbitrary vector spaces of dim u & n, respectively.

$$\begin{tikzcd}
M(n \times u) \arrow[rr, leftrightarrow, "(1)"] \arrow[dr, leftrightarrow, "(3)"'] & & \mathcal{L}(R_u, R_n) \arrow[dl, leftrightarrow, "(2)"] \\
& \mathcal{L}(V, W) &
\end{tikzcd}$$

Unfortunately, (2) will not be unique (as (1) was), so will get lots of one-to-one correspondences (3).

To get (2), recall facts about isomorphisms between $V \approx R_u$, $W \approx R_m$ [$\approx$: is isomorphic with].

To get isomorphism from V to R_u, we must pick a basis $a_1, \dots, a_u$ for V. If $x \in V$, x has the unique expression:

$$x = \sum_{i=1}^{u} \lambda_i a_i$$

Call λ_i coordinates of x relative to the basis $a_1, \dots, a_u$.

<u>Def</u>: $C(x) = (\lambda_1, \dots, \lambda_u)$ give a function $C: V \to R_u$. We proved C is an isomorphism – call it the <u>coordinate isomorphism</u> determined by this particular basis.

C has an inverse isomorphism $C^{-1}: R_u \to V$, $\ni$ $C^{-1}(\lambda_1, \dots, \lambda_u) = \sum \lambda_i a_i$
Note that C & C^{-1} depend upon the basis chosen, & in general will be different for different bases

<u>Notation</u>:

V & W are vector spaces of dimension u & m, respectively.
$a_1, \dots, a_u$ is a basis of V, coordinate isomorphism C
$a'_1, \dots, a'_m$ is a basis of W, " " C'
$T: V \to W$ is a linear transformation

We now associate with T a linear transformation $\tilde{T}: R_u \to R_m$ (2)

$$\begin{tikzcd} R_u \arrow[r, "\tilde{T}"] \arrow[d, "C^{-1}"'] & R_m \\ V \arrow[r, "T"'] & W \arrow[u, "C'"'] \end{tikzcd}$$

<u>Def</u>: $\tilde{T} = C' T C^{-1}$

Now (1). We know that $\tilde{T}$ is essentially just matrix multiplication of column vectors by a matrix $A = A_{\tilde{T}}$.

A is thus defined to be the matrix of T relative to the bases $a_1, \dots, a_u$ of V and $a'_1, \dots, a'_m$ of W

Note: This is a generalization of previous (1) association in the sense that if $T: R_u \to R_n$; then A_T = matrix of T relative to bases $(e_1, \dots, e_u)$ and $(e_1', \dots, e_u')$.

As before, Prove:

Theorem: If $a_1, \dots, a_u$ and $a_1', \dots, a_n'$ are bases of V and W, respectively, and $T: V \to W$ is a linear transformation; let $\Phi(T)$ be the matrix of T relevant to these bases. So $\Phi(T) \in \mathcal{M}(n \times u)$, and ~~thus~~ thusly, we have defined a function $\Phi: \mathcal{L}(V,W) \to \mathcal{M}(n \times u)$. This Φ is one-to-one and onto.

Proof: Φ is composition of two functions $\mathcal{L}(V,W) \xrightarrow{\Phi_2} \mathcal{L}(R_u, R_n) \xrightarrow{\Phi_1} \mathcal{M}(n \times u)$.

$$\left.\begin{array}{l} \Phi_2(T) = C'TC^{-1} = \tilde{T} \\ \Phi_1(\tilde{T}) = A_{\tilde{T}} \quad \text{[old definition] \& \underline{notation}} \end{array}\right\} \Phi = \Phi_1 \Phi_2$$

We know all about Φ_1 (one-to-one & onto); hence if Φ_2 is one-to-one & onto, Φ is one-to-one & onto $\Rightarrow$ isomorphism.

To prove that Φ_2 is one-to-one & onto, consider

$$\theta_2 : \mathcal{L}(R_u, R_n) \to \mathcal{L}(V,W)$$

$$\theta_2(S) = C^{-1\prime} S C$$

Claim – $\Phi_2 \theta_2$ = identity function of $\mathcal{L}(R_u, R_n)$
$\theta_2 \Phi_2$ = " " " $\mathcal{L}(V,W)$

Therefore, by Elementary Theorem 4 (on ~~sheet~~ notes),

θ_2 & Φ_2 are inverse function, one-to-one, & onto. Q.E.D.

Computations: (1) If we know $A = (\alpha_{ij})$ [$n \times u$ matrix], to get corresponding $T: V \to W$ (Φ^{-1}):

Defin: $A = C'TC^{-1}$ where we write A for T_A, hence:

$$C^{-1\prime} A C = C^{-1\prime} C'(T) C^{-1} C = I(T) I$$

Linear Transformation from V to W; The Matrix of a Transformation Relative to Bases; equivalence

$$\begin{tikzcd} R_u \arrow[r, "A"] & R_n \\ V \arrow[u, "C"] \arrow[r, "T"'] & W \arrow[u, "C'"'] \end{tikzcd}$$

A matrix acting by " multiplication

Def: $A \leftrightarrow T$ provided $A = C'TC^{-1}$

$T: V \to W$

The following conditions are logically equivalent to $A = C'TC^{-1}$

(1) $A = C'TC^{-1}$
(2) $C'^{-1}A = TC^{-1}$
(3) $C'^{-1}AC = T$
(4) $AC = C'T$

Note C & C' can be interpreted to be the linear transformations carrying the chosen bases to the natural bases.

Given $A = (\alpha_{ij})$, to get T (where $A \leftrightarrow T$), will use (3) above. This (3) says that for all $x \in V$, $T(x) = (C'^{-1}AC)(x)$

Compute: $C(x) = (\lambda_1, \ldots, \lambda_u)$ where $x = \sum_{j=1}^{u} \lambda_j a_j$

$$A(C(x)) = A(\lambda_1, \ldots, \lambda_u) = \begin{pmatrix} \alpha_{ij} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_u \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{u} \alpha_{1j} \lambda_j \\ \vdots \\ \sum_{j=1}^{u} \alpha_{nj} \lambda_j \end{pmatrix}$$

$$C'^{-1}(A(C(x))) = C'^{-1}\left(\sum_{j=1}^{u} \alpha_{1j}\lambda_j, \ldots, \sum_{j=1}^{u} \alpha_{nj}\lambda_j \right) = \sum_{i=1}^{n} \left(\sum_{j=1}^{u} \alpha_{ij} \lambda_j \right) a_i'$$

Result: If $x = \sum_{j=1}^{u} \lambda_j a_j$, then $T(x) = \sum_{i=1}^{n} \left(\sum_{j=1}^{u} \alpha_{ij} \lambda_j \right) a_i'$ or:

$$T\left(\sum_{j=1}^{u} \lambda_j a_j\right) = \sum_{i=1}^{n} \left(\sum_{j=1}^{u} \alpha_{ij} \lambda_j \right) a_i' \quad *$$

This last equation $(*)$ expresses the fact that $T\vec{c}' = \vec{c}''A$ since it says that T of the vector in V whose components are $\lambda_1, \ldots, \lambda_n$ is equal to the vector in W whose coordinates are the matrix multiple of $(\lambda_1, \ldots, \lambda_n)$ by (α_{ij}), the matrix of T.

Given T, to find the associated matrix $A = (\alpha_{ij})$:

By "given T", it is meant that $T(x)$ is known for every $x \in V$. Now T is characterized by the vectors $T(a_1), \ldots, T(a_n)$. That means that $A = (\alpha_{ij})$ must be determined by these vectors.

Claim: for $1 \le j \le n$,

$$T(a_j) = \sum_{i=1}^{n} \alpha_{ij} a_i' \qquad (**)$$

Proof in homework #8

Note reversal of indices

$\sum_j \alpha_{ij}\lambda_j$ is numbers (scalar)

then $\sum_i \alpha_{ij} a_i'$ is vector

Note: If the convention was changed to prevent reversal in $(**)$ [a vector], then they would be reversed in the scalars.

This shows how to get (α_{ij}): Express $T(a_j)$ as a linear combination of $a_1', \ldots, a_n'$. Then α_{ij} will be the i^{th} coefficient of $T(a_j)$ [relative to $a_1', \ldots, a_n'$].

Matrix Multiplication

Let $S: R_m \to R_n$ & $T: R_n \to R_p$ be linear transformations. S&T have matrices A_S $(n \times m)$ and A_T $(p \times n)$ associated with them. Since $TS: R_m \to R_p$ is determined by S&T, its matrix A_{TS} must be determined by the matrices A_S & A_T:

$$\left.\begin{array}{ll} S \leftrightarrow A_S = (\sigma_{ij}) & n \times m \\ T \leftrightarrow A_T = (\tau_{ij}) & p \times n \end{array}\right\} TS \leftrightarrow A_{TS} = (\alpha_{ij}) \quad p \times m$$

From before, we know that $TS(e_j)$ $[1 \le j \le m]$ is the j^{th} column of $(\alpha_{ij}) = (\alpha_{1j}, \alpha_{2j}, \ldots, \alpha_{pj}) = \sum_{i=1}^{p} \alpha_{ij} e_i''$

$\boxed{e_i \text{ in } R_m \text{ ; } e_i' \text{ in } R_n \text{ ; and } e_i'' \text{ in } R_p \; \boxed{\text{? notation}}}$

In order to determine A_{TS} in terms of (σ_{ij}) & (τ_{ij}), compute $(TS)(e_j)$ $[1 \le j \le m]$ two different ways

(1) $(TS)(e_j) = j^{th}$ column of $(\alpha_{ij}) = (\alpha_{1j}, \alpha_{2j}, \ldots, \alpha_{mj})$

$= \sum_{i=1}^{p} \alpha_{ij} e_i''$ (** for case where bases are ~~■~~ e's)

(2) $(TS)(e_j) = T(S(e_j)) = T\left(\sum_{k=1}^{n} \sigma_{kj} e_k'\right) = \sum_{k=1}^{n} \sigma_{kj} T(e_k')$

$\overset{*}{=} \sum_{k=1}^{n} \sigma_{kj}\left(\sum_{i=1}^{p} \tau_{ik} e_i''\right) = \sum_{i=1}^{p}\left(\sum_{k=1}^{n} \tau_{ik}\sigma_{kj}\right) e_i''$

→ But the last terms of (1) & (2) are equal, so:

$$\sum_{i=1}^{p} \alpha_{ij} e_i'' = \sum_{i=1}^{p} \sum_{k=1}^{n} \tau_{ik}\sigma_{kj} e_i''$$

But since a vector has a unique expression as a linear combination of $e_1'', \ldots, e_p''$; hence:

$$\alpha_{ij} = \sum_{k=1}^{n} \tau_{ik}\sigma_{kj} \qquad \begin{array}{l} 1 \le i \le p \\ 1 \le j \le m \end{array}$$

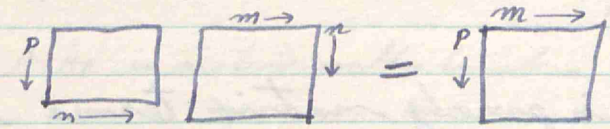
This is now taken as the <u>definition</u> of (τ_{ij}) times (σ_{ij})

$= (\tau_{ij})(\sigma_{ij}) \ne (\sigma)(\tau)$

So,... To find the entry in the i^{th} row, j^{th} column of the product:

Take the i^{th} row $(\tau_{i1}, \tau_{i2}, \dots, \tau_{in})$ of (τ_{ij}) and the j^{th} column $(\sigma_{1j}, \sigma_{2j}, \dots, \sigma_{nj})$ of (σ_{ij}). Multiply corresponding terms together $[\tau_{ik}\sigma_{kj}]$ and add.

$$\alpha_{ij} = \sum_{k=1}^{n} \tau_{ik}\sigma_{kj}$$



$$(p \times n)(n \times m) \longrightarrow (p \times m)$$

Example:

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

Note: $5 = (1)(1) + (2)(2) + (-1)(0)$; $-1 = (1)(1) + (2)(0) + (-1)(2)$

$1 = (1)(1) + (0)(2) + (1)(0)$; $3 = (1)(1) + (0)(0) + (1)(2)$

Hence: Given a basis $a_1, \dots, a_n$ for V and $a_1', \dots, a_n'$ for W, each linear transformation $T: V \to W$ has an $n \times n$ matrix $\Phi(T)$.

EQUIVALENCE

<u>Definition</u>: Two $n \times u$ matrices A & B are <u>equivalent</u> provided they both are associated with some one linear transformation $T: V \to W$ but relative to (possibly different) bases for V & W.

Notation: A eq B [$\Leftrightarrow$ + $\equiv$] A & B are equivalent.

<u>Two problems</u>:

(1) How to tell in purely matrix terms when A is equivalent to B.

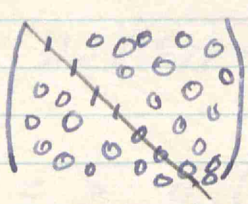
(2) Given A, what is the simplest matrix equivalent to it.

<u>Theorem</u> ①: If A & B are $n \times u$ matrices, A is equivalent to B if and only if there exists an $n \times n$ matrix P ~~and~~ of rank n, and a $u \times u$ matrix Q of rank u such that $B = PAQ$.

<u>Proof</u>: Homework #9

<u>Theorem</u> ②: Given an $n \times u$ matrix A, A is equivalent to an $n \times u$ matrix $D = (\delta_{ij}) \ni$

$$\delta = \begin{cases} 1 & \text{if } 1 \le i = j \le r \quad \text{where } r = \text{rank of } A. \\ 0 & \text{otherwise} \end{cases}$$

$D =$ 

Exactly r non-zero entries along <u>main diagonal</u> (45° line)

Examples
$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{matrix} 5 \times 3 \\ r = 3 \end{matrix} \qquad \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{matrix} 2 \times 5 \\ r = 2 \end{matrix}$$

<u>Facts</u>:

(1) If A is equivalent to B, and B is equivalent to C, then A is equivalent to C.

(2) If A is equivalent to B, then B is equivalent to A.

(3) Two matrices $(n \times u)$ A and B are equivalent if and only if they have the same rank; i.e. B is the matrix of $T_A: R_u \to R_n$ relative to

suitable bases.

Here will return to previous notation: T_A and A_T. A_T always relative to the unit basis. Note matrix multiplication has the following logically equivalent properties:

(1) $A_{TS} = A_T A_S$ → $A_T A_S$ is matrix multiplication as is T_{AB}

(2) $T_{AB} = T_A T_B$ → $T_A T_B$ is composition of functions as is A_{TS}

Proof of properties on p. (44) at bottom:

<u>(2)</u> is obvious from the definition: neither A or B has a preferred role in the symmetrical definition.

<u>(3) ⇒ (1)</u>: If A is equivalent to B [B equivalent to A], by (3) A is the matrix of T_B relative to suitable bases. Also, B is equivalent to C, so C is the matrix of T_B relative to some bases. Hence, A is equivalent to C.

<u>(3)</u>: Long & trivial proof by using facts on mimeo sheet and definitions over & over.

Assume A and B are equivalent. Then ∃ a linear transformation $T: V \to W$ and coordinate isomorphisms C, C', D, D' ∋ $T_A = C' T C^{-1}$ and $T_B = D' T D^{-1}$

$$\begin{tikzcd} R_n \arrow[r, "T_A"] & R_m \\ V \arrow[u, "C"] \arrow[r, "T"'] & W \arrow[u, "C'"'] \end{tikzcd} \qquad \begin{tikzcd} R_n \arrow[r, "T_A"] & R_m \\ V \arrow[u, "D"] \arrow[r, "T"'] & W \arrow[u, "D'"'] \end{tikzcd}$$

Solve first equation for T and substitute into second.

$$T = C'^{-1} T_A C \Rightarrow T_B = D' C'^{-1} T_A C D^{-1} = (D' C'^{-1})\, T_A\, (C D^{-1})$$

This would be the statement that B is the matrix of T_A relative to certain bases if it were in the form $T_B = E' T_A E^{-1}$ with E & E' coordinate isomorphisms.

Proof of this over:

$(D'^{-1}C')$: C' is an isomorphism and therefore, C'^{-1} is an isomorphism, & therefore, $D'C'^{-1}$ is an isomorphism from R_n to R_n. Let $E' = D'C'^{-1}$. Then E'^{-1} exists since E' is an isomorphism. (Every isomorphism is a coordinate isomorphism.)

Now, will show E' is the coordinate isomorphism of a certain basis for R_n:

Let $e_1, \dots, e_n$ be the natural basis, and $a_j = E'^{-1}(e_j)$. Then $E'(a_j) = E'(E'^{-1}(e_j)) = I(e_j) = e_j$. Hence, $E'(\sum \lambda_i a_i) = \sum \lambda_i E'(a_i') - \sum \lambda_i e_i = (\lambda_1, \dots, \lambda_n)$. Thus, E' is the coordinate isomorphism of the basis $a_1', \dots, a_n'$.

~~Similarly, a basis can be found~~

Similarly, a coordinate system $a_1, \dots, a_u$ can be found for R_u $\ni$ its coordinate isomorphism E has $E^{-1} = CD^{-1}$. <u>Q.E.D.</u>

Theorem ②: [c.f. p. 44] Every $n \times u$ matrix A is equivalent to $D = (\delta_{ij}) \ni$

$$\delta = \begin{cases} 1 & \text{if } 1 \le i = j \le r = \text{rank of the matrix } A. \\ 0 & \text{otherwise} \end{cases}$$

<u>Proof</u>:

Let A be an $n \times u$ matrix and consider $T_A : R_u \to R_n$. As in the proof that rank + nullity = u, let $a_{r+1}, \dots, a_u$ be a basis for the null space of T_A. Extend this to a basis $a_1, \dots, a_r, a_{r+1}, \dots, a_u$ for R_u. We know that $T_A(a_{r+1}) = \dots = T_A(a_u) = 0$; and $T_A(a_1), \dots, T_A(a_r)$ is a basis for the range of T_A. Call these vectors $a_1', \dots, a_r'$, $\ni T_A(a_i) = a_i'$. Since they are linearly independent, they can be extended to a basis $a_1', \dots, a_n'$ of R_n. Let D be the matrix of T_A relative to these bases. Calculate D by using (**) which says

$$T_A(a_j) = \sum_{i=1}^{n} \delta_{ij} a_i' \quad \text{where } (\delta_{ij}) \text{ is the matrix } D \qquad (1)$$

If $1 \le j \le r$, then $T(a_j) = a_j'$ (2)

Hence, the only way (1) & (2) can both hold is:

$$\delta_{ij} = 0 \text{ if } i \ne j \quad \& \quad \delta_{jj} = 1.$$

Hence δ_{ij} is the <u>Kronecker</u> δ.

If $r < j \leq n$, $T_A(a_j) = 0$ so $\delta_{ij} = 0$ for all i.

Combination shows D as claimed. D is the matrix of T_A, and hence, D & A are equivalent by definition because A is also a matrix of T_A (relative to the natural basis). So A & D are matrices of T_A & hence, equivalent.

A equivalent to B $\overset{(m \times n)}{\Leftrightarrow}$ A & B have same rank.

Proof: $\Rightarrow$] If A & B are equivalent, they are the matrices of a linear transformation $T: V \to W$ relative to suitable bases. In fact, can take T to be $T_A : R_n \to R_m$. By definition, rank A = rank T_A, and rank B = rank T_B but $T_B = C' T_A C^{-1}$ (where C & C' are coordinate isomorphisms) and can easily prove rank T_A = rank T_B because C and C' are isomorphisms. Hence,

$$\text{Rank } B = \text{rank } T_B = \text{rank } C' T_A C^{-1} = \text{rank } T_A = \text{rank } A.$$

$\Leftarrow$] If rank A = rank B = r; A and D are equivalent and B and D are equivalent where $D = (\delta_{ij})$. Hence A is equivalent to B.

Comments on method:

All of this theory is for abstract vector spaces and linear transformations. The vector space R_n is only an example — an important example since any finite-dimensional vector space is isomorphic to some R_n — but still only an example.

Matrices are not taken as a basic object of study. They appear when we consider the special vector space R_n. Once we understand the relationship between the general idea of a linear transformation and a matrix, we can translate information about linear transformations into information about matrices.

Hence, the question arises: Why not study R_n and matrices and then translate in inverse direction to get information about linear transformations and abstract vector spaces?

Abstract vector spaces occur in nature - with the vectors being one single object; e.g.,

(1) Set of all forces on a point mass
(2) Set of all solutions to a linear differential equation.
etc., etc., etc.

Rn doesn't occur naturally - Rn occurs when we introduce coordinate systems; i.e., pick a basis for a vector space and get ~~the~~ a coordinate isomorphism. But ~~no~~ there is no preferred coordinate system; so one can be chosen that best fits the problem.

Purpose of coordinates is to reduce natural solution to numerical problem whose answer can be computed. So in a typical scientific problem (Rn & matrices), coordinates are one stage of the solution and have no intrinsic significance.

(1) Sometimes a problem can be solved directly in abstract vector space terms without introducing Rn. This is particularly likely in case the problem has a yes or no answer instead of a number.

(2) Vector spaces and linear transformations generalize to the infinite dimensional case, but matrices do not.

(3) If interested only in computation and not their origin or application, then it pays to study matrices directly. Methods of efficient computation can be developed in special cases without the use of linear transformations.

(III) Linear Operators

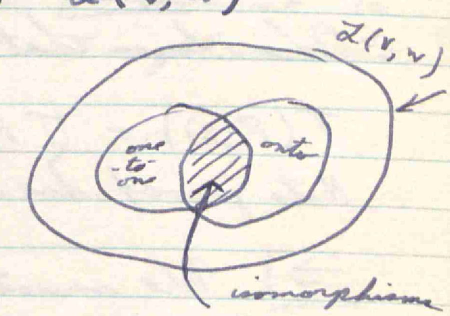
Change of terminology:

Def: A <u>linear operator on V</u> is simply a linear transformation T from vector space V into V, i.e., $T: V \to V$. We assume that V is an n-dimensional vector space with real or complex numbers as scalars.

Let $\mathcal{L}(V)$ be the set of all linear operators on V. From previous theorem, since $\dim V = \dim V$, a linear operator is one-to-one ~~if and only if~~ $\Leftrightarrow$ T is onto $\Leftrightarrow$ T is an isomorphism. A linear operator T which is an isomorphism is said to be <u>non-singular</u>.

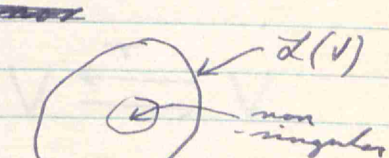
Note: There are four types of transformations in $\mathcal{L}(V, W)$

(1) isomorphisms (both one-to-one and onto)
(2) one-to-one, not onto
(3) onto, not one-to-one
(4) neither one-to-one or onto.



But in $\mathcal{L}(V)$ a linear operator is either ~~not~~

(1) non-singular
or (2) not non-singular (singular)



Linear operators on an n-dimensional vector space:

$$\boxed{\text{rank } T = n} \Longleftrightarrow \boxed{T \text{ onto}} \overset{①}{\Longleftrightarrow} \boxed{\begin{array}{l} T \text{ has a right inverse } R\text{, i.e.} \\ R \text{ operator } \ni TR = I \end{array}}$$

$$\Updownarrow$$

$$\boxed{\text{nullity of } T = 0} \Longleftrightarrow \boxed{T \text{ one-to-one}} \overset{②}{\Longleftrightarrow} \boxed{\begin{array}{l} T \text{ has a left inverse } L\text{, i.e.,} \\ L \text{ operator } \ni LT = I \end{array}}$$

$$\Updownarrow$$

$$\boxed{T \text{ is non-singular}} \Longleftrightarrow \boxed{T \text{ has inverse } T^{-1}}$$

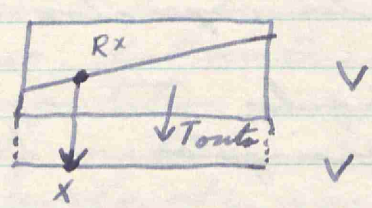
Proof of ①:

$\Leftarrow$ | If $x \in V$, we must show x is hit by something. We claim that $R(x)$ works.

$T(R(x)) = (TR)(x) = Ix = x$; hence, T is onto.
Note proof doesn't use linearity

$\Rightarrow$ | We must find an operator $R \ni T(Rx) = x$ for all $x \in V$.

$$V \underset{\longleftarrow}{\overset{T}{\longrightarrow}} V$$



Since T is onto, for each $j = 1, \dots, n$, there is a vector $b_j \ni T(b_j) = a_j$. Let R be the linear operator $\ni$ $R(a_j) = b_j$. We claim that $TR = I$. It suffices to prove it true for the basis, i.e., $(TR)(a_j) = I a_j$. But $(TR)(a_j) = T(R(a_j)) = T(b_j) = a_j = I(a_j)$.

Proof of ②:

Notice that ① and ② are actually true for any linear transformation $T: V \to W$, not just the linear operator $T: V \to V$. In the general case, T can have a right inverse but not a left inverse (hence not an inverse), ~~however in~~ and conversely, it can have a left inverse, but not a right inverse. However, in case of linear operators, we have:

$$R = L = T^{-1} \qquad \text{(notations as above)}$$

Right ~~of~~ or left inverse is an inverse.

Proof:

Suppose $TR = I$. Then T is onto, & hence, T is non-singular or T has an inverse T^{-1}. (An inverse is characterized by being simultaneously a left and right inverse.)

$$R = IR = (T^{-1}T)R = T^{-1}I = T^{-1}$$

Hence, $R = T^{-1}$

Proof similar for left inverse.

Analogy to proof: (real numbers)
Given: $t\gamma = 1$, then $\gamma \stackrel{?}{=} \frac{1}{t}$
Proof: $\gamma = 1 \times \gamma = \frac{t}{t} \times \gamma = \frac{1}{t} \times t\gamma = \frac{1}{t}$

Definition: A linear algebra L is a vector space furnished with another operation (called multiplication) which to vectors a, b, assigns a vector ab (called the product of $a + b$). Multiplication has following properties:

Left inverse: $(\lambda a + \mu b)c = \lambda ac + \mu bc$

Right inverse: $c(\lambda a + \mu b) = \lambda ca + \mu cb$

Associative Law: $(ab)c = a(bc)$

for all vectors a, b, c and all scalars λ, μ.

(Brief ~~notation~~ def.: Linear algebra is a vector space with a bilinear multiplication.

Def: If $ab = ba$ for all $a \in L$, $b \in L$, L is called a commutative algebra.

If there is a vector $e \in L \ni ex = xe = x$ for all $x \in V$, e is called a unit (or unit vector), and L is an algebra with a unit.

Example: R = real numbers = both vectors and scalars.
VA = usual addition
VM = SM = usual multiplication

Theorem: Let $\mathcal{L}(V)$ be the set of all linear operators on V, a finite dimensional vector space. $\mathcal{L}(V)$ is an algebra with a unit provided ~~must~~ operations are defined as follows:

SM: λT is the operator $\ni (\lambda T)(x) = \lambda(Tx)$ for all $x \in V$.

VA: $S+T$ is the operator $\ni (S+T)(x) = S(x) + T(x)$ for all $x \in V$.

VM: ST is the composition of functions, i.e., $(ST)(x) = S(T(x))$ for all $x \in V$.

Sample of proof: $(\lambda S + \mu T)U \overset{?}{=} \lambda(SU) + \mu(TU)$

This holds $\Leftrightarrow$ these two operators have the same value for every vector in V:

$$([\lambda S + \mu T][U])(x) = (\lambda S + \mu T)(U[x])$$

$$= (\lambda S)(U[x]) + (\mu T)(U[x]) = \lambda\{S[U(x)]\} + \mu\{T[U(x)]\}$$

$$= \lambda\{(SU)(x)\} + \mu\{(TU)(x)\} = (\lambda SU)(x) + (\mu TU)(x) = [\lambda(SU) + \mu(TU)](x).$$

<u>Theorem</u>: Let $\mathcal{M}(n\times n)$ be the set of all $n\times n$ matrices

Define: SM: $\lambda(\alpha_{ij}) = (\lambda\alpha_{ij})$

VA: $(\alpha_{ij}) + (\beta_{ij}) = (\alpha_{ij} + \beta_{ij})$

VM: $(\alpha_{ij})(\beta_{ij}) = \left(\sum_{k=1}^{n} \alpha_{ik}\beta_{kj}\right)$

Then $\mathcal{M}(n\times n)$ is an algebra with a unit.

Recall that usual one-to-one correspondence between $\mathcal{L}(R_n)$ and $\mathcal{M}(n\times n)$ holds.

Claim: This correspondence preserves algebra operations.

<u>Lemma</u>: $T_{\lambda A} = \lambda T_A$; $T_{A+B} = T_A + T_B$; $T_{AB} = T_A T_B$ and

$A_{\lambda T} = \lambda A_T$; $A_{S+T} = A_S + A_T$; $A_{ST} = A_S A_T$

Proof trivial but long.

Now we can easily prove that $\mathcal{M}(n\times n)$ is analgebra. Sampleofproof:

$(\lambda A + \mu B)C \overset{?}{=} \lambda AC + \mu BC$

<u>Proof:</u>

Since the function which carries a matrix A to its linear operator T_A is one-to-one; to prove two matrices are the same, it suffices to prove they have the same linear operator.

$$T_{(\lambda A+\mu B)C} = T_{(\lambda A+\mu B)} T_C = (\lambda T_A + \mu T_B) T_C \qquad \text{by lemma.}$$

We know linear operators are algebra; hence

$$= \lambda (T_A T_C) + \mu (T_B T_C) = \lambda T_{AC} + \mu T_{BC} = T_{\lambda AC} + T_{\mu BC} = T_{(\lambda AC + \mu BC)}$$

Claim: The unit of $\mathcal{M}(n\times n)$ is the matrix associated with the identity transformation I [unit of $\mathcal{L}(R_n)$], A_I

By definition: j^{th} column of A_I is $I(e_j) = e_j$

$$A_I = \begin{pmatrix} 1\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0 \\ 0\,0\,1\,0\,0 \\ \text{-- -- -- --} \end{pmatrix} \qquad A_I = (\delta_{ij}) \quad ; \; I = (\delta_{ij})$$

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

Note: $\mathcal{M}(n \times n)$ and hence, $\mathcal{L}(V)$ are not commutative.

Example: $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$

$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$

Non-Singular Matrices

Def: An $n \times n$ matrix A is non-singular provided the corresponding linear operator R_n is non-singular.

A is non-singular $\Leftrightarrow$ T_A is non-singular $\Leftrightarrow$ rank $T_A = n$

$\Leftrightarrow$ rank $A = n$ $\Leftrightarrow$ row rank $A = n$ (computable)

If A is a non-singular $n \times n$ matrix, define A-inverse (A^{-1}) to be the matrix corresponding to T_A^{-1} of the linear transformation T_A corresponding to A.

Def: A matrix has an inverse $\Leftrightarrow$ it is non-singular.

$$\begin{array}{ccc} A & \longrightarrow & T_A \\ & & \downarrow \\ A^{-1} & \longleftarrow & T_{A^{-1}} \end{array} \qquad \text{or } T_{A^{-1}} = T_A^{-1}$$

Properties:

① A is non-singular $\Rightarrow A^{-1}$ exists & is non-singular
$\Rightarrow (A^{-1})^{-1} = A$
$\Rightarrow A A^{-1} = A^{-1} A = I$

② B and C are $n \times n$ matrices $\ni BC = I$
$\Rightarrow B$ and C are non-singular
$\Rightarrow B = C^{-1}$; $C = B^{-1}$
$\Rightarrow BC = CB = I$ (i.e., commutativity)

Proof of ① in Homework

Proof of ②:

$$BC = I \Rightarrow T_{BC} = I \Rightarrow T_B T_C = I \Rightarrow T_B \text{ and } T_C$$

are non-singular and inverses [proved previously]

$\Rightarrow$ B and C are nonsingular and inverses.

To see last step, notice that by definition of inverse, $T_C^{-1} = T_{C^{-1}}$, so $T_B = T_{C^{-1}}$; hence, $B = C^{-1}$ due to the one-to-one correspondence.

Computations of A^{-1}

① Elementary row operations can be accomplished by left matrix-multiplication by a suitable matrix. The result of applying a certain elementary row operation to A is EA, where E is that matrix resulting from the application of that certain row operation to I, the identity matrix.

The matrix representing an elementary row operation is called an elementary matrix.

$$I = \begin{pmatrix} 1&0&0&0&0 \\ 0&1&0&0&0 \\ 0&0&1&0&0 \\ 0&0&0&1&0 \\ 0&0&0&0&1 \end{pmatrix}$$

Elementary Row Operation $\longleftrightarrow$ Elementary Matrix

Interchanging rows. [(1↔2)] $\longleftrightarrow$ $\begin{pmatrix} 0&1&0&0&0 \\ 1&0&0&0&0 \\ 0&0&1&0&0 \\ 0&0&0&1&0 \\ 0&0&0&0&1 \end{pmatrix}$

Multiplying one row by some constant. [row 3 by λ] $\longleftrightarrow$ $\begin{pmatrix} 1&0&0&0&0 \\ 0&1&0&0&0 \\ 0&0&\lambda&0&0 \\ 0&0&0&1&0 \\ 0&0&0&0&1 \end{pmatrix}$

Add one row to a different row. [add 2 to 4] $\longleftrightarrow$ $\begin{pmatrix} 1&0&0&0&0 \\ 0&1&0&0&0 \\ 0&0&1&0&0 \\ 0&1&0&1&0 \\ 0&0&0&0&1 \end{pmatrix}$

(II) An $n \times n$ matrix A is non-singular $\Leftrightarrow$ its super-echeloned form is I

Proof:

Let A^* be super-echeloned A. Then A is non-singular $\Leftrightarrow$ rank $A = n \Leftrightarrow$ rank $A^* = n$ $\Leftrightarrow$ A^* contains no zero rows $\Leftrightarrow A^* = I$.

This last step can be seen by recalling that a super-echeloned matrix has ~~only~~ all zeros except the leading entry of a row; since A^* is $n \times n$ and has n non-zero rows, it is I.

(III) $BC = I = CB$.

Proof: Put (I) and (II) together: Suppose A is a non-singular $n \times n$ matrix. Derive a method of computing A^{-1} as follows: We know every matrix can be super-echeloned by application of the elementary row operations. Hence, by (I) and (II), $\exists$ elementary matrices $E_1, E_2, \dots E_K$ $\ni$ $E_K E_{K-1} \cdots E_2, E_1 A = A^* = I$. Multiplying both sides by A^{-1}:

$\underline{\underline{E_K E_{K-1} \cdots E_1 I = A^{-1}}}$. Therefore, to compute A^{-1}, super-echelon it, getting I, and simultaneously apply the same elementary row operations on I; result is A^{-1}.

NOTE: $(E_K E_{K-1} \cdots E_1) A = I$; $(E_K E_{K-1} \cdots E_1) I = A^{-1}$

so $(E_K E_{K-1} \cdots E_1)(E_K E_{K-1} \cdots E_1) A = A^{-1} = (E_K E_{K-1} \cdots E_1)^2 A$

$A^{-1} = E_K E_{K-1} \cdots E_1$

EXAMPLE:

$$\left\{\begin{pmatrix} 1 & 3 & -2 \\ 1 & 0 & 1 \\ 0 & 3 & -1 \end{pmatrix} \atop \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right\} \rightarrow \left\{\begin{pmatrix} 1 & 3 & -2 \\ 0 & -3 & 3 \\ 0 & 3 & -1 \end{pmatrix} \atop \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right\} \rightarrow \left\{\begin{pmatrix} 1 & 3 & -2 \\ 0 & -3 & 3 \\ 0 & 0 & 2 \end{pmatrix} \atop \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 1 \end{pmatrix}\right\} \rightarrow$$

$$\left\{\begin{pmatrix} 1 & 3 & -2 \\ 0 & +1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \atop \begin{pmatrix} 1 & 0 & 0 \\ 1/3 & -1/3 & 0 \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix}\right\} \leftrightarrow \left\{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \atop \begin{pmatrix} 1/2 & 1/2 & -1/2 \\ -1/6 & 1/6 & 1/2 \\ -1/2 & 1/2 & 1/2 \end{pmatrix}\right\} \leftrightarrow \begin{pmatrix} 1 & 3 & -2 \\ 1 & 0 & 1 \\ 0 & 3 & -1 \end{pmatrix}\begin{pmatrix} 1/2 & 1/2 & -1/2 \\ -1/6 & 1/6 & 1/2 \\ -1/2 & 1/2 & 1/2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

If A is non-singular, $\exists$ elementary matrices $E_1, E_2, \ldots, E_k \ni E_k E_{k-1} \cdots E_2 E_1 = A^{-1}$

Every inverse is a product of elementary matrices. Every non-singular matrix, however, is an inverse (viz., the inverse of its inverse). Hence;

<u>Every non-singular matrix is a product of elementary matrices.</u>

<u>Similarity of Matrices</u>:

<u>Notation</u>: $\begin{cases} A \sim B \Leftrightarrow A \;\&\; B \text{ are similar} \\ A \text{ eq } B \Leftrightarrow A \;\&\; B \text{ are equivalent} \end{cases}$

The matrix of a linear transformation $T: V \to W$ relative to bases $a_1, \ldots, a_n$ for V and $a_1', \ldots, a_u'$ for W has been previously defined. This led to the idea of equivalence of matrices [i.e., two matrices representing the same linear transformation.]. The following two problems were solved:

(1) Find a purely matrix criterion that two $n \times u$ matrices be equivalent. [$B = PAQ$; P, Q non-singular]

(2) Find the simplest matrix equivalent to any given matrix. [A eq (δ_{ij})]. The solution of this also involves the problem: Given two matrices (both $n \times u$), are they equivalent? [A eq $B \Leftrightarrow$ rank A = rank B; compute by super-echeloning.]

$$\left.\begin{array}{ll} A \text{ eq } D_r & (r = \text{rank } A) \\ B \text{ eq } D_s & (s = \text{rank } B) \end{array}\right\} \quad D_r = \begin{pmatrix} 1 & & & & 0 \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ 0 & & & & 0 \end{pmatrix} = (\delta_{ij}) \text{ where } \begin{array}{l} 1 \le i \le n \\ 1 \le j \le u \\ 0 = i = j > r \end{array}$$

D_r eq $D_s \Leftrightarrow D_r = D_s \Leftrightarrow r = s$; hence A eq $B \Leftrightarrow$ rank A = rank B.

Now, repeat this work for the special case $V = W$ and $a_i = a_i'$, i.e., for linear operators on V, using only one basis.

To be considered:

(1) Special cases of equivalence

(2) Analogy to equivalence

(3) Will turn out to be more difficult and require new methods.

<u>Fundamental Definition</u>: Let T be a linear operator on V; $a_1, \dots, a_n$ be a basis for V; and C be the coordinate isomorphism $\ni C: V \to R_n$.

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "C"'] & V \arrow[d, "C"] \\ R_n \arrow[r, "T_A"'] & R_n \end{tikzcd}$$

CTC^{-1} is an operator on R_n, so $\exists$ a matrix $A \ni T_A = CTC^{-1}$; A is called the matrix of T relative to $a_1, \dots, a_n$. It follows from the previous work that this association of a linear operator and a matrix, for a fixed basis, gives a one-to-one correspondence between $\mathcal{L}(V) \leftrightarrow \mathcal{M}(n \times n)$. Also, $A = (\alpha_{ij})$ is the matrix of T relative to $a_1, \dots, a_n \Leftrightarrow T_A = CTC^{-1} \Leftrightarrow^{(*)}$

<u>(*) & (**) refer to pages 40 & 41</u>

$$\overset{*}{\Leftrightarrow} T\Big(\sum_{i=1}^{n} \lambda_i a_i\Big) = \sum_{j=1}^{n}\Big(\sum_{i=1}^{n} \alpha_{ij}\lambda_i\Big) a_j \overset{(**)}{\Leftrightarrow} T(a_j) = \sum_{i=1}^{n} \alpha_{ij} a_i$$

<u>Analogy of equivalence [2 bases] in single basis case</u>:

<u>Def.</u>: Two $n \times n$ matrices are similar provided they are the matrices of the same linear operator relative to possibly different bases (one basis per matrix).

<u>Note</u>: Similarity is a relationship applicable only to square matrices; while equivalence applies to any rectangular matrices (of the same dimension). But even in ~~[illegible]~~ considering equivalence in the square ($n \times n$) case, it is not as close a relationship as similarity. Since given an $n \times n$ matrix A, the matrices equivalent to A are, as proved, <u>all</u> matrices which represent $T_A: R_n \to R_n$ relative to any choice of the bases $a_1, \dots, a_n$; $a_1', \dots, a_n'$. In the same way, it can be shown that the matrices similar to A are all matrices which represent T_A relative to any one basis $a_1, \dots, a_n$, i.e., $a_j = a_j'$. Thus, all similar matrices are equivalent but the converse does not necessarily hold. The greater freedom in the case of equivalence suggests that there will be more matrices equivalent to A than similar.

Notation: $A \sim B \Leftrightarrow A$ & B are similar

Matrix criterion for similarity [analogous to that for equivalence]

Two $n \times n$ matrices, A and B, are similar $\Leftrightarrow$ $\exists$ a non-singular matrix, P, $\ni$

$$B = P^{-1}AP$$

Proof: [$\Rightarrow$] Let $T: V \to V$.

$A \sim B$ means that $\exists$ a basis $a_1, \ldots, a_n$ (with the corresponding coordinate isomorphism C_a) and a basis $b_1, \ldots, b_n$ (" " " " " C_b), $\ni$ A is the matrix of T relative to $a_1, \ldots, a_n$ and B is the matrix of T relative to $b_1, \ldots, b_n$.

Proof $\Leftarrow$ on page 58.

Properties of similarity:

(1) $A \sim B \Rightarrow A \text{ eq } B$

(2) $A \sim B \Rightarrow A \sim C$ if $B \sim C$

(3) $A \sim B \Leftrightarrow B$ is the matrix of T_A relative to a suitable basis, or, A is the matrix of T_B relative to a suitable basis.

Proofs:

(1) See p. 56

(3) $\Rightarrow$ (2) Just as in the case of equivalence.

(3) [$\Leftarrow$]. This is obvious since if A is the matrix of T_A relative to natural basis, so if B is the matrix of T_A relative to some basis, then $A \sim B$ by definition.

(3) [$\Rightarrow$] This is exactly the same as for equivalence except that $V = W$; $a_j = a_j'$ (i.e., ~~one~~ only one basis).

$A \sim B \Rightarrow$ the same transformation diagram ~~holds~~ as for equivalence holds except for above changes. ($V=W$, $C=C'$, $D=D'$) ($C=C'$ & $D=D'$ because for similarity the bases must be the same: $a_j = a_j'$; $b_j = b_j'$.)

$$\begin{tikzcd}
R_n \arrow[r, "T_A"] & R_n \\
V \arrow[u, "C"] \arrow[r, "T"] \arrow[d, "D"'] & V \arrow[u, "C"'] \arrow[d, "D"] \\
R_n \arrow[r, "T_B"'] \arrow[uu, bend left=60, "E"] & R_n \arrow[uu, bend right=60, "E"']
\end{tikzcd}$$

Here, $C =$ coordinate isomorphism of a_i
$D =$ coordinate isomorphism of b_i $\Big\}$ $1 \le i \le n$

$E = CD^{-1}$; i.e., diagram is commutative.

E is an isomorphism; D^{-1} and C are isomorphisms; hence, E is a coordinate isomorphism of some basis. But since the diagram is commutative, $T_B = E^{-1} T_A E$. This means that B is the matrix of T_A relative to some basis, and therefore, $B \sim A$ $\boxed{(3)\ p.\ 57}$

Matrix criterion for similarity: $A \sim B \Leftrightarrow B = PAP^{-1}$; P non-singular (p. 57) proof $\Rightarrow$

<u>Proof:</u> ($\Leftarrow$) If $P^{-1}AP = B$: $T_B = T_{P^{-1}AP} \Rightarrow T_{P^{-1}} T_A T_P = T_B \Rightarrow$ $T_B = T_P^{-1} T_A T_P$ [since T_P is non singular because P is non-singular]. Hence, $T_P = C$, the coordinate isomorphism of some basis; but $T_B = C^{-1} T_A C$ says that B is the matrix of T_A relative to this basis. Therefore, $A \sim B$. <u>(3) p. 57</u>.

<u>Proof:</u> ($\Rightarrow$) Since $A \sim B$, use above transformation diagram:

$T_B = E^{-1} T_A E$; Since $E: R_n \to R_n$, $E = T_P$ where P is a matrix. P is non-singular because E is. Then

$T_B = T_P^{-1} T_A T_P = T_{P^{-1}} T_A T_P = T_{P^{-1}AP}$ & $\therefore$ $B = P^{-1}AP$.

Extra information is gained here: a formula for the matrix P, analogous to $(**)$ on p. 41, not involving T.

<u>Terminology</u>: $E = CD^{-1}$ = change of coordinate <u>operator</u> (from b to a)

P = matrix of E relative to natural basis
= change of coordinate matrix from b to a.

— — — — — — — — — —

<u>Lemma</u>: If P = change of coordinate matrix from b to a,

$$P = (\pi_{ij}), \text{ where } b_j = \sum_{i=1}^{n} \pi_{ij} a_i \qquad (***)$$

<u>Proof</u>: By definition, the j^{th} column of P is $T_P(e_j) = E(e_j)$

$= C(D^{-1}(e_j))$, which can be computed.

<u>Roughly speaking</u>,

T	takes	a vector x	to	a vector Tx.
A	"	"a" coordinates of x	"	"a" coordinates of Tx
B	"	"b" coordinates of x	"	"b" coordinates of Tx
P	"	"b" coordinates of x	"	"a" coordinates of x

Thus, $B = P^{-1}AP$
since b coord $x \xrightarrow{B} b$ coord Tx is same as

b coord $x \xrightarrow{P} a$ coord $x \xrightarrow{A} a$ coord $Tx \xrightarrow{P^{-1}} b$ coord Tx.

There are two ways of finding the matrix B of a linear operator relative to a given basis $b_1, \dots, b_n$.

(1) Use $(**)$: $T(b_j) = \sum_{i=1}^{n} \beta_{ij} b_i$ where $B = (\beta_{ij})$

(2) From another known matrix A of T, relative to another basis $a_1, \dots, a_n$.

$$\begin{tikzcd}
\boxed{\text{coords of } x \text{ rel. to } a_1,\dots,a_n} \arrow[r, "A"] & \boxed{\text{coords of } Tx \text{ rel to } a_1,\dots,a_n} \\
\boxed{\text{coords of } x \text{ rel to } b_1,\dots,b_n} \arrow[u, "P"] \arrow[r, "B"'] & \boxed{\text{coords of } Tx \text{ rel to } b_1,\dots,b_n} \arrow[u, "P"]
\end{tikzcd}$$

Let P be the matrix of change of coordinates from b-coordinates to a-coordinates; i.e.,

$$b_j = \sum_{i=1}^{n} \pi_{ij} a_i \qquad (***)$$

<u>Def:</u> Trace $(\alpha_{ij}) = \sum_{i=1}^{n} \alpha_{ii}$ where $(\alpha_{ij}) = A$ is $n \times n$.

A matrix is singular $\Leftrightarrow$ 0 is an eigenvalue.

C J Whitehead

<u>Theorem</u> $n \times u$ matrices A and B are equivalent $\Longleftrightarrow$ There exists an $n \times n$ matrix P of rank n and a $u \times u$ matrix Q of rank u such that $B = PAQ$

<u>Proof $\Rightarrow$</u> If A and B are equivalent, we proved that B is the matrix of $T_A: R_u \to R_n$ relative to suitable bases. By definition this means that if C & C' are the coordinate isomorphisms of these bases, then $T_B = C' T_A C^{-1}$. Thus

$$B = A_{(T_B)} = A_{(C' T_A C^{-1})} = A_{C'} A_{(T_A)} A_{C^{-1}} = A_{C'} A A_{C^{-1}}$$

where as usual A_T means the matrix of T relative to the usual bases (of e_j's.) But C' and C^{-1} are isomorphisms so their matrices have ranks n & u respectively. So if we let $P = A_{C'}$ and $Q = A_{C^{-1}}$ the proof is complete.

<u>Proof $\Leftarrow$</u>

Suppose $B = PAQ$ with P & Q as given in the statement of the theorem. These matrices determine linear transformations as usual and we claim $T_B = T_P T_A T_Q$. To prove this note that

$$PAQ = B = A_{T_B} \;\&\; A_{T_P \cdot T_A T_Q} = A_{T_P} A_{T_A} A_{T_Q} = PAQ$$

that is, T_B & $T_P T_A T_Q$ have the <u>same</u> matrix and hence must be equal (the correspondence: linear transf $\longleftrightarrow$ matrix is one-to-one.) Since P & Q have ranks n & u, so do T_P & T_Q, i.e. they are isomorphisms, and T_Q^{-1} is too. Thus T_P and T_Q^{-1} are, as proved in class, the coordinate isomorphisms of a suitable choice of bases. Thus $T_B = T_P T_A (T_Q^{-1})^{-1}$ says B is the matrix of T_A relative to these bases. And A is the matrix of T_A relative to the usual bases, so A & B are equivalent.

Functions

A **set** A is any collection of objects, called **elements** of A. The notation $a \in A$ means a is an element of A. A **function** $\phi: A \to B$ from set A to set B is a rule which assigns to each $a \in A$ an element $\phi(a) \in B$. A function ϕ is **one-to-one** provided $\phi(a) = \phi(b) \Rightarrow a = b$. A function ϕ is **onto** provided for each $b \in B$ there is an $a \in A$ such that $\phi(a) = b$. If $\phi: A \to B$ and $\psi: B \to C$ are functions, denote by $\psi\phi: A \to C$ the function such that $(\psi\phi)(a) = \psi(\phi(a))$ for all $a \in A$. $\psi\phi$ called **composition** of ϕ and ψ.

Def: If ϕ is both one-to-one & onto, ϕ is an **isomorphism**.

Some elementary theorems

(1) If $\phi: A \to B$ is one-to-one and onto, there is a unique function $\phi^{-1}: B \to A$ such that $\phi^{-1}(b) = a \iff \phi(a) = b$. ϕ^{-1} is called the inverse of ϕ.

(2) If ϕ is one-to-one and onto, then ϕ^{-1} is one-to-one and onto, and $(\phi^{-1})^{-1} = \phi$.

(3) If $\phi: A \to B$ and $\psi: B \to C$ are both one-to-one and both onto, then $\psi\phi$ is one-to-one and onto, and $(\psi\phi)^{-1} = \phi^{-1}\psi^{-1}$.

(4) If $\phi: A \to B$ and $\theta: B \to A$ are functions such that $\theta\phi = I_A$ (identity function of A) and $\phi\theta = I_B$ (identity function of B) then both ϕ and θ are one-to-one and onto, and are inverses, i.e. $\phi = \theta^{-1}$ and $\theta = \phi^{-1}$.

~~$A \sim B,\ B \sim C \Rightarrow A \sim C$~~

~~To prove this need another stuff:~~

[To avoid confusion go back to old notation

T_A = lin tr dtd by A]

(3) $A \sim B \iff A$ is the matrix of

$T_B : R_n \to R_n$ rel to suitable

bases (not unc e's) of R_n & R_n

<u>Proof.</u> $A \sim B$ means by defn

that $\exists\ T : V \to W$ & coord isoms

$C, C', D, D' \ni \quad T_A = C'\, T\, C^{-1}$

$T_B = D'\, T\, D^{-1}$

$$\begin{tikzcd} R_n \arrow[r, "T_A"] & R_n \\ V \arrow[u, "C"] \arrow[r, "T"'] & W \arrow[u, "C'"'] \end{tikzcd}$$

Solve 2nd for T & subst in 1st

$T = D'^{-1}\, T_B\, D$

$T_A = (C'\, D'^{-1})\, T_B\, (D\, C^{-1})$

Interpret this.

I

1. Vector Spaces
2. Examples
3. Linear Independence & Span
4. Linear transformations
5. One-to-one and onto
6. Computations in R_n

II Linear Transformation & Matrices

1. Linear transformations from R_n to R_n
2. Linear equations
3. Composition of a linear transformations & matrix multiplication
4. Linear transformation from L to L'; Matrices relative to bases; equivalence.
5. Linear transformation from L to L.

III Linear Operators

1. Non-singularity; non-singular operators
2. Algebraic properties of $\mathcal{L}(V)$ and matrices. ; algebra
3. Non-singular matrices.
4. Similarity of matrices

The Diagram for the Equivalence of Matrices

Two $n \times n$ matrices A & B are equivalent provided they represent the same linear transformation — this involves the following:

Linear transf. $T : V \to W$

A, the matrix of T relative to the bases $a_1, \ldots, a_n$ of V (coordinate isomorphism C) and $a'_1, \ldots, a'_n$ of W (coord. isom. C')

B, the matrix of T relative to the bases $b_1, \ldots, b_n$ of V (coord isom. D) and $b'_1, \ldots, b'_n$ of W (coord isom D')

These objects and the relations between them are summarized in the following diagram:

$$\begin{tikzcd}
R_n \arrow[r, "T_A"] & R_m \\
V \arrow[u, "C"] \arrow[d, "D"'] \arrow[r, "T"] & W \arrow[u, "C'"'] \arrow[d, "D'"] \\
R_n \arrow[r, "T_B"] \arrow[uu, bend left=60, "E"] & R_n \arrow[uu, bend right=60, "E'"']
\end{tikzcd}$$

where $E = CD^{-1}$ and $E' = C'D'^{-1}$

These two equations plus the equations $T_A = C'TC^{-1}$ and $T_B = D'TD^{-1}$ (which express the fact that A & B are matrices of T) have the following consequence:

<u>the diagram is commutative</u>, that is, any two ways of going from one vector space to another are the same (using any indicated linear transformation or, the inverse of any of the 'vertical' linear transfs — which are all isomorphisms)

For example: $E'T_B = C'TD^{-1} =$

$E'D'TC^{-1}E = T_AE$

III-5 Determinants

References: Thomas *Calculus*, Birkhoff & Maclane.

The determinant of a linear operator is a number, of considerable algebraic and geometric significance. We define first the determinant of an $n \times n$ matrices. To do this we need the notion of permutation. A *permutation* of $1, 2, \ldots, n$ is simply a rearrangement of these integers in a definite order $j_1, j_2, \ldots, j_n$. There are $n!$ such permutations.

Example The permutations of $1, 2, 3$ are:

123, 132, 213, 231, 312, 321

The operation of reversing two adjacent integers in a permutation is called a *transposition*. Every permutation of $1, 2, \ldots, n$ can be gotten by a finite number of transpositions but there are in general lots of ways of doing this

Example

$$\begin{tikzcd}
123 \arrow[r] \arrow[d] & 132 \arrow[r] & 312 \\
213 \arrow[r] & 231 \arrow[r] & 321 \arrow[u]
\end{tikzcd}$$

Theorem If a permutation can be reached from $1, 2, \ldots, n$ by an even number of transpositions in one way, then any way will consist of an even number of transpositions. (Hence any way odd $\Rightarrow$ all ways odd.)

(For proof see Birkhoff & MacLane) 2

Thus set of all permutations splits into two subsets: <u>even permutations</u> & <u>odd permutations</u> depending on evenness or oddness of number of transpositions necessary to reach them from $1, 2, \ldots, n$.

<u>Example</u> $n=3$

Even	Odd
123	132
231	213
312	321

Define $\sigma(j_1 j_2 \cdots j_n) = \begin{cases} +1 & \text{if } j_1 j_2 \cdots j_n \text{ is even} \\ -1 & \text{if } j_1 j_2 \cdots j_n \text{ is odd} \end{cases}$

Now we can define the determinant of a matrix

<u>Defn.</u> If $A = (\alpha_{ij})$ is an $n \times n$ matrix, the determinant of A is the number:

$$\det A = \sum \sigma(j_1 j_2 \cdots j_n)\, \alpha_{1 j_1} \alpha_{2 j_2} \cdots \alpha_{n j_n}$$

where the sum is over all $n!$ of the permutations $j_1 j_2 \cdots j_n$ of $1, 2, \ldots, n$.

This formula says: form all products of n entries of A — no two in same row or same column. Then $\det A = \sum \pm (\text{products})$ where sign is det'd by permutations.

Example

$$\det\begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix} = \begin{matrix} +\alpha_{11}\alpha_{22}\alpha_{33} + \alpha_{12}\alpha_{23}\alpha_{31} + \alpha_{13}\alpha_{21}\alpha_{32} \\ -\alpha_{11}\alpha_{23}\alpha_{32} - \alpha_{12}\alpha_{21}\alpha_{33} - \alpha_{13}\alpha_{22}\alpha_{31} \end{matrix}$$

Thus this definition generalizes the usual rules for 2×2 and 3×3 matrices.

The main properties of determinants are:

(1) A non-singular $\Longleftrightarrow$ $\det A \neq 0$

(2) $\det(AB) = \det A \cdot \det B$

(3) Expansion along row or column - say the k^{th} row:

$$\det A = \sum_{j=1}^{n} (-1)^{k+j} \alpha_{kj} \det A_{kj}$$

where A_{kj} is the $(n-1) \times (n-1)$ matrix obtained by striking out the k^{th} row and j^{th} column of A.

We shall prove the first two of these - the test for non-singularity and the product rule. For the third, which is an efficient way to compute determinants, see Birkhoff & MacLane. (Note how (3) works in the example at the top of this page.)

Our definition of determinant is in terms of the standard notation (α_{ij}) for matrix, where i is row index, j column index. If a matrix is not in standard notation, ignore any indices it may have, and interpret α_{ij} in the definition as the entry in i^{th} row and j^{th} column.

Example $\det \begin{pmatrix} \alpha_{71} & \alpha_{43} \\ \alpha_{12} & \alpha_{11} \end{pmatrix} = \sigma(1,2)\,\alpha_{71}\alpha_{11} + \sigma(2,1)\,\alpha_{43}\alpha_{12}$

$$= \alpha_{71}\alpha_{11} - \alpha_{43}\alpha_{12}.$$

Properties ① and ② follow easily from this special case of ②:

Lemma If E is an elementary matrix, then $\det EA = \det E \det A$.

Proof. Check for each of the three types of elementary matrices. Switches of rows and scalar multiplication of rows are easy, so we give the proof only for the addition of one row to another – say adding second row to the first. In this case:

$$E = \begin{pmatrix} \begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array} & & & & 0 \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ 0 & & & & 1 \end{pmatrix}$$ and thus $\det E = 1$

$$EA = \begin{pmatrix} \alpha_{11}+\alpha_{21} & \alpha_{12}+\alpha_{22} & \cdots & \alpha_{1n}+\alpha_{2n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ & \text{etc.} & & \end{pmatrix}$$

Thus by definition

$$\det EA = \sum \sigma(j_1 j_2 \cdots j_n)\,(\alpha_{1j_1} + \alpha_{2j_1})\,\alpha_{2j_2} \cdots \alpha_{nj_n}$$

$$= \underbrace{\sum \sigma(j_1 j_2 \cdots j_n)\,\alpha_{1j_1}\alpha_{2j_2} \cdots \alpha_{nj_n}} + \sum \sigma(j_1 j_2 \cdots j_n)\,\alpha_{2j_1}\alpha_{2j_2} \cdots \alpha_{nj_n}$$

The first sum is just $\det A$, so to complete the proof we need only prove the second sum is zero. The set of permutations of $1, 2, \ldots, n$ may be arranged into pairs which differ only by a transposition of their first two numbers. Corresponding to each such pair are two terms in the second sum:

$$\sigma(r, s, j_3 \cdots j_n)\,\alpha_{2r}\alpha_{2s}\alpha_{3j_3} \cdots \alpha_{nj_n} + \sigma(s, r, j_3, \ldots, j_n)\,\alpha_{2s}\alpha_{2r}\alpha_{3j_3} \cdots \alpha_{nj_n}$$

But $\alpha_{2r}\alpha_{2s}\alpha_{3j_3} \cdots \alpha_{nj_n} = \alpha_{2s}\alpha_{2r}\alpha_{3j_3} \cdots \alpha_{nj_n}$

while $\sigma(r, s, j_3, \ldots, j_n) = -\sigma(s, r, j_3, \ldots, j_n)$

So the terms in the second sum all cancel out in pairs and it is zero.

It is now easy to prove 1 and 2

1 Theorem A non-singular $\Leftrightarrow \det A \neq 0$

Proof. The preceding proof showed that performing any one of the elementary operations on a matrix changes its determinant by multiplication by a non-zero number (± 1, λ, $+1$ in the three cases). Thus if A^* is a superechelon of A, $\det A^* = \alpha \det A$ with $\alpha \neq 0$ because get A^* by a series of elementary row operations.

If A nonsingular then $A^* = I$ so

$$\det A = \frac{1}{\alpha} \det I = \frac{1}{\alpha} \cdot 1 \neq 0.$$

If A singular A^* contains a zero row, hence $\det A^* = 0$ because every term in the formula for its determinant contains an $\alpha_{ij} = 0$. Thus

$$\det A = \frac{1}{\alpha} \det A^* = \frac{1}{\alpha} \cdot 0 = 0.$$

(2) Theorem. $\det(AB) = \det A \det B$

Proof. Consider two cases:

If A singular then $\det A = 0$ and also AB singular (because A singular $\Rightarrow$ T_A not onto $\Rightarrow$ $T_{AB} = T_A T_B$ not onto $\Rightarrow$ AB singular) hence $\det AB = 0$, and the theorem is trivially true

If A nonsingular then A is a product $E_1 E_2 \cdots E_k$ of elementary matrices, and:

$\det AB = \det(E_1 E_2 \cdots E_k B) =$
$\det E_1 \det E_2 \cdots \det E_k \det B =$
$\det(E_1 E_2 \cdots E_k) \det B = \det A \cdot \det B$

(The second & third equalities follow by induction using the preceding lemma.)

Corollary 1 If A nonsingular, $\det(A^{-1}) = \dfrac{1}{\det A}$

Proof. Take determinant of both sides of $AA^{-1} = I$.

Corollary 2 Similar matrices have the same determinant.

Proof. If A and B are similar, then $B = PAP^{-1}$ (with P nonsingular) hence $\det B = \det P \det A \det P^{-1}$
$= \dfrac{\det P}{\det P} \det A = \det A$.

This last corollary shows that determinants are really properties of linear operators, for all the matrices of an operator T (being similar) have the same determinant, which we define to be $\det T$.

The determinant of an operator is its "coefficient of expansion." For example consider $T: R_3 \to R_3$. If X is a region in R_3 which has volume ρ, then it can be proved that T transforms X into a region X' with volume $\rho \det T$.

Alternative notation: $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$ instead of $\det\begin{pmatrix} a & b \\ c & d \end{pmatrix}$

III-4)

In II-5 we proved that every $n \times n$ matrix is equivalent to a matrix of a certain standard simple form — namely, all zeroes except for some 1's at the beginning of the main diagonal. (This led to a computable criterion for equivalence: A equivalent B $\Longleftrightarrow$ same rank)

There is no analogous proof in the case of similarity, for in that proof (of Thm 2 in II-5) we used our freedom to pick two bases, picking one well adapted to nullspace of T_A, another well adapted to the range of T_A. But for similarity we can only pick one basis, so the method fails.

There is however a theorem analagous to Thm 2 where the standard simple form is the so-called Jordan canonical form (and there are other theorems for other forms) but we will not have time to prove them: they require new methods.

Instead we consider a related problem of no less importance, namely:

Take as the simple type of matrixes the diagonal matrices, i.e. those whose entries are zero except (possibly) on the main diagonal, and ask:

1. Which matrices are similar to a diagonal matrix? (not all are)

2. If the $n \times n$ matrix A is similar to a diagonal matrix (= if A is diagonalizable) how do we compute P and L (diagonal) such that $P^{-1}AP = L$?

III-6 Eigenvalues and Eigenvectors.

By (**) we know that a linear operator T has a diagonal matrix relative to basis $a_1, \ldots, a_n$ if $T(a_j)$ is just a scalar multiple of a_j (for every j). Our aim is to find conditions for a given $n \times n$ matrix A to be similar to a diagonal matrix, so we will find the following concepts useful:

Defn. Let T be a linear operator on V. If λ is a number and x a non-zero vector of V such that $Tx = \lambda x$, then:

λ is an eigenvalue of T

x is an eigenvector of T belonging to λ

The eigenvalues and eigenvectors of an $n \times n$ matrix A are defined to be those of T_A.

Obviously every eigenvector of T belongs to one and only one eigenvalue, but an eigenvalue may have more-than-one linearly independent eigenvectors. An operator may have no eigenvalues and eigenvectors, for example $T =$ rotation of the plane through $90°$. Since we are interested in computing eigenvectors and values we first consider linear operators T_A on R_n, then later the case of an arbitrary operator on V.

Theorem A Let A be an $n \times n$ matrix,

$L = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \ \lambda_n \end{pmatrix}$ a diagonal $n \times n$ matrix. Then the following are logically equivalent:

(1) A is similar to L.

(2) there is a basis $p_1, p_2, \dots, p_n$ of R_n composed of eigenvectors of T_A whose eigenvalues are $\lambda_1, \lambda_2, \dots, \lambda_n$.

(3) $P^{-1}AP = L$ where P is a non-singular $n \times n$ matrix whose j^{th} column is p_j.

Among other things, this important theorem answers the question: when is a square matrix diagonalizable? (i.e. similar to a diagonal matrix) It is diagonalizable $\Longleftrightarrow$ its eigenvectors span (for then and only then can we get a basis of eigenvectors.)

Now we turn from theory to computations. How can we compute whether a matrix A is diagonalizable and (if it is) find P & L as in (3)? What is needed is this second, and equally important theorem:

(1) $\Rightarrow$ (2)

$A \sim L \Rightarrow L$ is the matrix of T_A relative to suitable basis $P_1, \dots, P_n$

$\Rightarrow^{**} T_A(P_j) = \sum_{i=1}^{n} \lambda_{ij} P_i$ where $L = (\lambda_{ij})$

But L is diagonal $\Rightarrow \lambda_{ij} = \begin{cases} \lambda_{ii} & \text{when } i = j \\ 0 & \text{otherwise} \end{cases}$

Hence $**$ reduces to $T_A(P_j) = \lambda_{jj} P_j$

(over)

(2) ⇒ (3): This follows from the general (indirect) method of computing the matrix of T_A relative to $P_1, \dots, P_n$ when we know its matrix relative to $e_1, \dots, e_n$.

However, can give direct proof:

$AP = PL$ where P is the matrix whose j^{th} column is P_j

$$AP = (A)\begin{pmatrix} | & & | \\ P_1 & \cdots & P_n \\ | & & | \end{pmatrix} = \begin{pmatrix} | & | & & | \\ \lambda_1 P_1 & \lambda_2 P_2 & \cdots & \lambda_n P_n \\ | & | & & | \end{pmatrix} \text{ since } (A)\begin{pmatrix} | \\ P_j \\ | \end{pmatrix} = \begin{pmatrix} | \\ \lambda_j P_j \\ | \end{pmatrix}$$

$$PL = \begin{pmatrix} | & | & & | \\ P_1 & P_2 & \cdots & P_n \\ | & | & & | \end{pmatrix}\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & 0 \\ & 0 & \ddots & \\ & & & \lambda_n \end{pmatrix} = \begin{pmatrix} | & & | \\ \lambda_1 P_1 & \cdots & \lambda_n P_n \\ | & & | \end{pmatrix} \quad \text{hence } AP = PL$$

P is non singular since column rank $= n \Rightarrow P^{-1}$ exists

$$P^{-1}AP = P^{-1}PL = IL = L$$

---

(3) ⇒ (1) Previously proved: matrix criterion for similarity.

<u>Theorem B</u> Let A be an $n \times n$ matrix, λ_0 a number. Then:

λ_0 is an eigenvalue of A

$$\Longleftrightarrow$$

there is a non-zero (eigen) vector $x_0 \in R_n$ such that $A x_0 = \lambda_0 x_0 = \lambda_0 I x_0 \Leftrightarrow A x_0 - \lambda_0 I x_0 = 0$

$$\Longleftrightarrow$$

the equation $(A - \lambda_0 I)x = 0$ has a non-trivial solution

$$\Longleftrightarrow$$

$A - \lambda_0 I$ is singular

$$\Longleftrightarrow$$

$$\det(A - \lambda_0 I) = 0$$

Above λ_0 is one number, below λ is a variable:

<u>Defn.</u> If A is an $n \times n$ matrix, then $\det(A - \lambda I)$ is, by definition of determinant, a polynomial in λ of degree n and is called the <u>characteristic polynomial</u> of A. The equation $\det(A - \lambda I) = 0$ is called the <u>characteristic</u> <u>equation</u> of A.

Using this terminology we get from Thm B:

<u>Corollary</u> The eigenvalues of A are the roots of its characteristic polynomial, i.e. the solutions of its characteristic equation (hence there are at most n) If λ_0 is an eigenvalue, the eigenvectors belonging to λ_0 are the non-trivial solutions of the equation $(A-\lambda_0 I)x = 0$.

Put the preceding facts together we have

<u>Method of Computing</u>

(1) whether a given $n \times n$ matrix is diagonalizable, and if so

(2) matrices P and L (diagonal) such that $P^{-1}AP = L$.

<u>Step 1</u> "Expand" $\det(A-\lambda I)$, using main fact 3 about determinants, to get characteristic polynomial $c_n\lambda^n + c_{n-1}\lambda^{n-1} + \dots + c_1\lambda + c_0$

<u>Step 2</u> Solve the char. equation obtaining the distinct eigenvalues $\lambda_1, \lambda_2, \dots \lambda_k$ of A. (We emphasize that eigenvalues are <u>scalars</u>, so if we are using the real numbers as the scalars of our vector spaces, only the <u>real</u> roots of the char. polynomial are eigenvalues.) k may be 0, i.e. there may be no eigenvalues.

i.e. all different

<u>Step 3</u> For each λ_i in $\lambda_1, \dots, \lambda_k$: Solve the equation $(A - \lambda_i I)x = 0$. [The solution subspace of the equation is called the <u>eigenspace</u> of λ_i, because it consists of 0 and all the eigenvectors belonging to λ_i. The dimension m_i of the eigenspace of λ_i is called the <u>geometric multiplicity</u> of λ_i, i.e. m_i = nullity $A - \lambda_i I$: m_i is the max. number of linearly indep eigenvectors belonging to λ_i] By "solve" we mean, as usual, find a basis for the solution subspace = eigenspace.

Now put all these basis eigenvector for all the λ_i's together — there will be $\sum_{i=1}^{k} m_i$ of them and they are independent (proved below) Thus they span (& A is diagonalizable) $\iff n = \sum_{i=1}^{k} m_i$ In words:

$n \times n$ A is diagonalizable $\iff$ the sum of the geometric multiplicities of its (distinct) eigenvalues is n.

<u>Step 4</u> If A is diagonalizable, there are n basis eigenvectors forming a basis for R_n Write them down in order $P_1, P_2, \dots, P_n$ so that the m_1 basis eigenvectors of λ_1 come first, then the

~~Proof of Lemma (1)~~

~~Suppose $P_1, \dots, P_k$ are dependent. $\exists$ a smallest integer $r \ni P_1, \dots, P_r$ are linearly independent but $P_1, \dots, P_{r+1}$ are dependent. $r > 0$ because $P_1 \neq 0$ and any one non-zero vector is linearly independent. Hence P_{r+1} can be expressed as a linear combination $\sum_{i=1}^{r} \mu_i P_i$. Since $P_1, \dots, P_r$ are independent, $\mu_1, \dots, \mu_r$ are uniquely determined. Applying T_A to the above linear combination gives $T(P_{r+1}) = \sum_{i=1}^{r} \mu_i T P_i = \lambda_{r+1} P_{r+1} = \sum_{i=1}^{r} \mu_i \lambda_i P_i$~~

~~$\lambda_{r+1} \neq 0$ since the P_i's $(1 \le i \le r)$ are linearly independent.~~

μ_i all not zero?

Proof of Lemma (1)

Suppose $P_1, \dots, P_k$ are dependent. Then $\exists$ a smallest integer $r \ni P_1, \dots, P_r$ are linearly dependent. $r > 1$, because any one non-zero vector is linearly independent. Set $\sum_{i=1}^{r} \mu_i P_i = 0$. Applying T_A to both sides: $0 = \sum_{i=1}^{r} \mu_i T P_i = \sum_{i=1}^{r} \mu_i \lambda_i P_i$. Multiplying the first combination by $(-\lambda_r)$ gives $0 = \sum_{i=1}^{r} (-\lambda_r) \mu_i P_i$. Addition of this to the second combination $\sum \mu_i \lambda_i P_i$ gives $\sum_{i=1}^{r} (\lambda_r - \lambda_i) \mu_i P_i = 0 = \sum_{i=1}^{r-1} (\lambda_r - \lambda_i) \mu_i P_i$. But the λ_i's are distinct and no μ_i $(1 \le i \le r) = 0$ for otherwise $P_1, \dots, P_r$ is not a minimal dependent set. Hence $\sum_{i=1}^{r-1} (\lambda_r - \lambda_i) \mu_i P_i = 0 \Rightarrow P_1, \dots, P_{r-1}$ are dependent but this contradicts the assumption that $P_1, \dots, P_{r-1}$ is minimal. Hence $P_1, \dots, P_k$ are linearly independent.

Proof of Lemma (2): Assume they are not independent. Then $\sum_{\text{all } ij} \mu_{ij} P_{ij} = 0$ with not all $\mu_{ij} = 0$. Let $P_i = \sum$ all terms $\mu_{ij} P_{ij}$ in this linear combination involving the eigenvectors of λ_i. Rewrite as $P_1 + \dots + P_k = 0$. Since (1) not all $\mu_{ij} = 0$ and (2) eigenvectors $P_{i1}, \dots, P_{ir_i}$ are independent, not all $P_1, \dots, P_k = 0$. Hence the non-zero ones are dependent; but the non-zero ones are eigenvectors of some of the original λ_i's. But this contradicts Lemma (1). So they are independent.

m_2 eigenvectors (basis) of λ_2, etc. Let P be the matrix whose j^{th} column is p_j. Let L be

$$\begin{pmatrix} \overbrace{\lambda_1 \; \lambda_1 \cdots \lambda_1}^{m_1} & & & 0 \\ & \overbrace{\lambda_2 \cdots \lambda_2}^{m_2} & & \\ & & \ddots & \\ 0 & & & \overbrace{\lambda_k \cdots \lambda_k}^{m_k} \end{pmatrix}$$

Then $P^{-1}AP = L$.

<u>Lemma 1</u> If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of A and p_j is an eigenvector belonging to λ_j ($j = 1, \ldots, k$) then the vectors $p_1, \ldots, p_k$ are linearly independent.

Briefly: eigenvectors belonging to distinct eigenvalues are linearly independent. For the procedure above we need a more complicated form of this fact

<u>Lemma 2</u> If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of A, and (for each $j = 1, \ldots, k$) $p_{j1}, p_{j2}, \ldots, p_{jr_j}$ is a linearly independent set of eigenvectors belonging to λ_j, then all p_{ij}'s are linearly independent.

[illegible] is

We now point out a few significant features of the preceding method. First, note that in addition to eigenvalues the characteristic polynomial contains ~~of~~ information about its matrix

A matrix is sing $\Leftrightarrow$ 0 is an eigenvalue

Lemma 3 If $c_n\lambda^n + c_{n-1}\lambda^{n-1} + \dots + c_1\lambda + c_0$ is the char. poly. of a square matrix A, then:

Wed 1-4

1) $c_n = (-1)^n$ where n is the "size" of A, i.e. A is $n \times n$

2) $c_{n-1} = (-1)^{n+1} \operatorname{tr} A^{\vee}$

3) $c_0 = \det A$

We have seen that a matrix may have no eigenvalues, but

Lemma 4 An $n \times n$ matrix with n odd has at least one eigenvalue.

(because its char. poly. has odd degree hence has a root.)

Like any root of a polynomial, an eigenvalue λ_0 has an *algebraic multiplicity* a_0, namely the largest integer such that $(\lambda - \lambda_0)^{a_0}$ divides the char. poly. without remainder. It can be shown that: geometric mult. $m_0 \leq$ algebraic mult a_0, but for the problem of diagonalizability, a_0 is not very important, because we can have $m_0 < a_0$. For example, let:

$$A = \begin{pmatrix} \gamma & 0 & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & \gamma \end{pmatrix}, \quad B = \begin{pmatrix} \gamma & 1 & 0 \\ 0 & \gamma & 1 \\ 0 & 0 & \gamma \end{pmatrix} \quad \text{where } \gamma \neq 0$$

A and B have the same char. poly.: $(\gamma - \lambda)^3$ So for both γ is an eigenvalue of alg. mult 3. But for A, γ has geom. mult 3 since e_1, e_2, e_3 are eigenvectors, while for B, γ has geom. mult 1 since e_1 (& its non-zero multiples) is the only eigenvector. Thus B is not diagonalizable.

$^{\vee}$ $\operatorname{tr} A$, the trace of A, is by defn $\sum_{i=1}^{n} a_{ii}$, the sum of the entries on the main diagonal.

while A is itself diagonal.

Thus, in general, we cannot tell from its char. poly. whether a matrix A is diagonalizable, even though the char. poly. contains such information as the size of A, $\operatorname{tr} A$, $\det A$, and all the eigenvalues of A with their algebraic multiplicities. However there is a special case when we can tell:

<u>Lemma 5</u> If the char. poly of an $n \times n$ matrix A can be factored into n <u>distinct</u> linear factors, i.e. if A has n distinct eigenvalues, then A is diagonalizable.

<u>Proof.</u> By lemma 1 the n eigenvectors in R_n are indep. — hence span.

When the scalars of our vector spaces are the complex numbers, the fundamental theorem of algebra says that the char. poly of a complex $n \times n$ matrix always has n (complex) roots, counting algebraic multiplicities, i.e. the sum of the algebraic multiplicities is n for every $n \times n$ complex matrix. ✓ (Thus lemma 4 is valuable only in the case of real scalars.) Thus also:

<u>Lemma 6</u> Every square complex matrix has at least one eigenvalue (and eigenvector).

Here for the first time is a difference between real and complex scalars. Since the real numbers α may also be considered to be the complex numbers $\alpha + i0$ a real matrix may also be considered to be a complex matrix, and in general its properties in the two cases will be very different. For example, let

---

✓ This does not mean every complex matrix is diagonalizable for the geom. mults may not add up to n. The matrix B on preceding page is not diagonalizable using complex scalars.

Proof of Lemma 7 $\quad T: V \to W$; $T_A: R_n \to R_n$

$$Tx = \lambda x \quad (x \neq 0) \; ; \; T = C^{-1} T_A C$$

$$\Leftrightarrow (C^{-1} T_A C)\, x = \lambda x \Leftrightarrow (C C^{-1})\, T_A C\, x = C \lambda x$$

$$\Leftrightarrow T_A (Cx) = \lambda (Cx).$$

$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ In both cases the charpoly of
A is λ^2+1. Considered as a real matrix then, A
has no eigenvalues, so, of course, is not diagonalizable. But
considered as a complex matrix it has eigenvalues $+i$
and $-i$ so by lemma 5 it is diagonalizable, and using
our previous methods (valid for complex as well as real scalars)
we can go on and compute the complex eigenvectors of A
and the complex matrix P such that $P^{-1}AP = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$.
Roughly speaking a real matrix is more likely
to be diagonalizable when considered as a complex
matrix than when considered for real scalars only. For
a complete and well-written treatment of the diagonal-
ization problem for complex vector spaces, see Halmos
Finite-dimensional Vector Spaces

So far we have worked only with matrices.
To apply the preceding results to linear operators
we need only relate the eigenvalues & vectors of an
operator to the eigenvalues and vectors of its matrix
relative to some basis.

Lemma 7 Let T be a linear operator on V, let
$a_1, \dots, a_n$ be a basis for V with coord. isom. $C: V \to R^n$
and let A be the matrix of T rel. to this basis. Then
the eigenvalues of T and A are the same, and
$x \in V$ is an eigenvector of T $\Longleftrightarrow$ its coords $C(x)$
form an eigenvector of A.

It follows that similar matrices have the

Proof of Lemma 8 :

If $A \sim B$, $B = P^{-1}AP$. Then characteristic polynomial of $P^{-1}AP$ is $\det(P^{-1}AP - \lambda I) = \det(P^{-1}AP - \lambda P^{-1}IP) = \det(P^{-1}AP - P^{-1}(\lambda I)P)$

$= \det\left[P^{-1}(A-\lambda I)P\right] = \det(A-\lambda I) =$ characteristic polynomial of A

→ Lemma 2 in determinant notes

same eigenvalues and their eigenvectors differ only by the change-of-coords isomorphism. Not only are their eigenvalues the same but more than that:

<u>Lemma 8</u> Similar matrices have the same characteristic polynomial.

Thus by lemma 3 similar matrices have the same trace$^{\Downarrow}$ and — known before — determinant. The true meaning of lemma 8 is that the char. poly. is really a property of a linear operator T since all matrices of T (rel. to different bases) have the same char. poly. Thus the same can be said for trace, determinant, eigenvalues, and (except for coord. isoms) eigenvectors. Diagonalizability too is a property of linear operators for if we define T to be diagonalizable provided there is a basis of V composed of eigenvectors of T, then all matrices of a diagonalizable operator will be diagonalizable in the previous sense.

$\Downarrow$ A consequence of this is the fact the trace of a diagonalizable matrix is the sum of its eigenvalues (each taken as many times as its geometric multiplicity).