Linear Algebra

20-Oct-09

Lecture Notes - 1

When asked the question "what is a vector", we tend to describe them as pts. in an n-dim. Euclidean space. In other words, "usually" a vector means a Euclidean vector. ~~Such vectors~~ We are familiar with such vectors and some basic operations on them are:

(i) Vectorial addition (component-wise addition) denoted by '+'

(ii) Scalar multiplication (component-wise ~~scaling~~ scaling) denoted by '·'

eg Take $\mathbb{R}^2$ and let $v_1 = (x_1, y_1)$, $V_2 = (x_2, y_2)$ and let $\alpha \in \mathbb{R}$

then $v_1 + v_2 = (x_1 + x_2, y_1 + y_2) \in \mathbb{R}^2$

$$\alpha \cdot v_1 = (\alpha x_1, \alpha y_1) \in \mathbb{R}^2$$

Note that these operations take in Euclidean vectors of dimension 'n' and give out Euclidean vectors of the same dimension. Hence ~~these~~ any set of vectors, of a particular dim., are closed under these two operations. In other words, +, · qualify to be "valid" binary and unary operators ~~on~~ on the set of all vectors (of a given dim.)

We ~~leave~~ differ analysis of other complicated operators like dot (inner) product and cross (outer) product (which do not produce vectors of same dimension as output) to a later stage.

It is easy to verify that the following 8 (fundamental) properties (axioms) are satisfied in case of Euclidean vectors together with the above mentioned +, · operators:

Let set of all vectors of a fixed dim. be represented by V

① Commutative law for $+$: $v_1 + v_2 = v_2 + v_1 \quad \forall v_1, v_2 \in V$

② Associative law for $+$: $(v_1 + v_2) + v_3 = v_1 + (v_2 + v_3) \quad \forall v_1, v_2, v_3 \in V$

3) Existence of $+$ identity: $\exists\, \omega$ such that $v_1 + \omega = v_1 \quad \forall v_1 \in V$

(ω is the zero vector of appropriate size)

④ Existence of $+$ inverse: for each $v_1 \in V$, $\exists$ a $\omega_1 \in V \ni v_1 + \omega_1 = \omega$

$+$ identity

$(-v_1$ is $\omega_1)$

(one can also show that $+$ identity and $+$ inverse are unique)

⑤ Associative law for $\cdot$: $\alpha, \beta \in \mathbb{R} \quad v \in V \Rightarrow (\alpha\beta) \cdot v = \alpha \cdot (\beta \cdot v)$

(multiplication of two scalars)

⑥ Identity for $\cdot$: $1 \cdot v = v \quad \forall v \in V$

Distributive laws: ⑦ $(\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v \quad \forall \alpha, \beta \in \mathbb{R} \quad \forall v \in V$

(sum of two scalars)

⑧ $\alpha \cdot (v_1 + v_2) = \alpha \cdot v_1 + \alpha \cdot v_2 \quad \forall \alpha \in \mathbb{R} \quad \forall v_1, v_2 \in V$

Now, we abstract the notion of vectors saying that any set of elements together with appropriately defined operation $+, \cdot$ qualifies to be a called a vector if the above 8 fundamental axioms are satisfied.

Here is the formal definition:

<u>Definition</u>: Let V be a non-empty set. Let $+$, $\cdot$ be a binary and unary operators defined on V (i.e. V is closed under $+$ and $\cdot$). The unary operator ~~is~~ $\cdot$ is such that it takes as input an element in V and a scalar in $\mathbb{R}$ and gives an element in V. The triplet $\mathcal{V}=(V,+,\cdot)$ is called as <u>a vector space</u> if the above mentioned 8 axioms are satisfied by it. Further, if $\mathcal{V}$ is a vector space, then elements in V are called as <u>vectors</u>.

~~Now without assuming any particular form for V, +, ·~~

Now, let $\mathcal{V}=(V,+,\cdot)$ be a vector space. Without assuming anything further about the nature of $V,+,\cdot$, one can prove the following from first principles (i.e. using 8 axioms alone):

(i) Commutativity & Associativity for $+$ together give that the symbols $v_1+v_2+v_3$ "makes sense". i.e. no need to put brackets, also addition in any order gives the same result. For eg:

$$\underbrace{(v_1+v_2)+v_3 \underset{\text{asso.}}{=} v_1+(v_2+v_3)}_{\text{comm.}} = (v_2+v_1)+v_3 \underset{\text{Comm.}}{=} v_3+(v_2+v_1)$$

$$\underset{\text{asso.}}{=} (v_3+v_2)+v_1 = \dots \text{ so on.}$$

By induction we can also make this statement for sum of n vectors and simply write $v_1+v_2+\dots+v_n$ (any bracketing and ordering is fine)

(ii) + identity is unique (see prop. 1.2 in pdf)
Hence we represent it by the symbol '0'.
(abuse of notation)

(iii) + inverse is unique (for each vector) (see prop. 1.3 in pdf)
Hence we represent + inverse of v by symbol $-v$

(iv) $$\underbrace{\underbrace{(\alpha\beta)\cdot v = \alpha\cdot(\beta\cdot v)}_{\text{asso.}} = (\beta\alpha)}_{\text{comm. for product of numbers}}\cdot v = \underbrace{\beta\cdot(\alpha\cdot v)}_{\text{asso.}}$$ → again abuse of notation

(v) $$(\alpha+\beta)\cdot(v_1+v_2) = \underbrace{\alpha\cdot(v_1+v_2)}_{\text{dist. ⑦}} + \underbrace{\beta\cdot(v_1+v_2)}_{\text{dist. ⑧}} = \alpha\cdot v_1 + \alpha\cdot v_2 + \beta\cdot v_1 + \beta\cdot v_2$$
$$= \underbrace{(\alpha+\beta)\cdot v_1}_{\text{dist ⑧}} + (\alpha+\beta)\cdot v_2$$

||| by extensions for m_1 vectors & m_2 numbers.

(vii) $0\cdot v = 0 \quad \forall\, v \in V$ (see prop. 1.4 in pdf)
0 → number zero; 0 → + identity

(viii) $\alpha\cdot 0 = 0 \quad \forall\, \alpha \in \mathbb{R}$ (see prop. 1.5 in pdf)
α → number; 0 → + identity; 0 → + identity

(ix) $-1\cdot v = -v \quad \forall\, v \in \mathbb{R}$ (see prop. 1.6 in pdf)
-1 → number -1; v → vector; $-v$ → its add. inverse

(x) $\alpha\cdot v = 0 \iff \alpha = 0$ or $v = 0$ (see your assign.)

Let give some eg. of vector spaces:

eg1 Consider V to be be set of all ~~vect~~ points in $\mathbb{R}^n$ (n is given)
Let + be "usual" vectorial addition & $\cdot$ be "usual" scalar mult.

$V = (\mathbb{R}^n, +, \cdot)$ is a vector space.

(This is actually how we obtained 8 axioms so obviously they are satisfied)

eg2 Consider V to be set of all sequences of real numbers.
+ is "usual" term-wise sum
$\cdot$ is "usual" term-wise scaling i.e.

$v_1 \in V \Rightarrow v_1 = (x_1, x_2, x_3, \dots, x_n, \dots)$

$v_2 \in V \Rightarrow v_2 = (y_1, y_2, y_3, \dots, y_n, \dots)$

$v_1 + v_2 \equiv (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n, \dots) \in V$

$\alpha \cdot v_1 \equiv (\alpha x_1, \alpha x_2, \dots, \alpha x_n, \dots) \in V$

It is easy to verify that $(V, +, \cdot)$ is a vector space.

~~The~~ Here identity element is sequence of all zeros.

eg3 Consider V to be set of all $m \times n$ matrices (m, n are given)
+ is matrix addition, $\cdot$ is "usual" multiplication of scalar and matrix

Again $(V, +, \cdot)$ is a Vector space.

Here identity element is a zero matrix.

eg4 Consider V to be set of all real valued functions on $\mathbb{R}$.
+ is "usual" sum of functions (Point-wise sum), $\cdot$ is "usual" point-wise scaling

$f \in V \Rightarrow \cancel{f}\ f(x) \in \mathbb{R} \quad \forall x \in \mathbb{R}$
$g \in V \Rightarrow g(x) \in \mathbb{R} \quad \forall x \in \mathbb{R}$

$$(f+g)(x) \equiv \underbrace{f(x)+g(x)}_{\in \mathbb{R}} \quad \forall x \in \mathbb{R} \Rightarrow f+g \in V$$

$$(\alpha \cdot f)(x) = \underbrace{\alpha f(x)}_{\in \mathbb{R}} \quad \forall x \in \mathbb{R}, \alpha \in \mathbb{R} \Rightarrow \alpha \cdot f \in V$$

Again $(V, +, \cdot)$ is a Vector space.

eg. Consider the set of all polynomial functions on $\mathbb{R}$ (call it V').
$V' \subseteq V$. It is easy to see that $(V', +, \cdot)$ is again a vector space (but is "contained" in $(V, +, \cdot)$ as $V' \subseteq V$).

set of all Polynomial functions on $\mathbb{R}$ — set of all real functions on $\mathbb{R}$

Consider V'' the set of all polynomial functions (on $\mathbb{R}$) with degree m (m is given). $(V'', +, \cdot)$ is not a vector space because if we add two polynomials of degree m, we might end up with polynomial degree $\leq m$.

Again, V''' the set of all poly. func. with degree $\leq m$ forms a vector space i.e. $(V''', +, \cdot)$ is indeed a Vector space.

eg5 Consider the set of all r.v.s (call it V). '+' is usual way to add two r.v.s, '$\cdot$' is usual way of scaling a r.v. We know from prob. theory that + & $\cdot$ both give a r.v again. Again ~~by similar steps for details~~, we can show that $(V, +, \cdot)$ is indeed a Vector space. Of course you need to assume that r.v.s are from some probability space.

Now consider V' the subset of r.v.s with mean zero. It is easy to see that $(V', +, \cdot)$ is again a vector space.

Ill by V'' the subset of r.v.s with finite second ~~data~~ moment also form a vector space.

~~Note the~~ following points are notable from the above examples:

(i) In case of eg 1 a "vector" means Euclidean vector
eg 2 " " a sequence of reals
eg 3 " " a matrix
eg 4 " " a function
eg 5 " " a r.v !

This should remove the notion that Euclidean vectors are the ONLY kind of vectors

(ii) +Identity element is : eg 1 $\longrightarrow$ "zero vector"
eg 2 $\longrightarrow$ "sequence of all zeros"
eg 3 $\longrightarrow$ zero matrix
eg 4 $\longrightarrow$ constant zero function
eg 5 $\longrightarrow$ degenerate r.v which takes value 0 with prob. 1.

Note that we denote each of these by '0' (symbol)

(iii) In many cases, subsets of the entire set of vectors itself formed a vector space! eg. collection of all polynomial functions is itself a vector space.

Such vector spaces lying insides other vector spaces are known as Sub-spaces.

~~Sub-spaces~~

(iv) All the eg. given above are eg. of ~~...~~ what are known as "real vector spaces" i.e. vector spaces which consider "scalars" as real numbers. (Recall that $\cdot$ takes a scalar, which in eg. of or defn. till now, always was a real number, and a vector to give a vector).
Now one can generalize this and define vector spaces which consider any other "Field" as the set of scalars. For eg. one can consider vector spaces over complex numbers or in other words complex vector spaces. In this course, we will consider only real vector spaces.

Sub-spaces

Let us now formally define sub-spaces:

Defn.: Given a vector space $V = (V, +, \cdot)$, consider the triplet $V' = (V', +, \cdot)$ where $V' \subseteq V$. If V' itself is a vector space, then V' is called a subspace in the vector space V.

Since $+, \cdot$ satisfy the 8 axioms over the entire set of vectors V, they will obviously (inheritance) be satisfied on the subset of vectors V'. However the only thing that needs to be checked is whether $+, \cdot$ remain to be valid operators on V' i.e. whether V' is closed under $+$ and $\cdot$.

Hence here is an equivalent defn. of subspace:

Defn Given a vector space $\mathbb{V} = (V, +, \cdot)$ and a subset $V' \subseteq V$, the triplet $\mathbb{V}' = (V', +, \cdot)$ is known as a subspace iff:

(i) V' is closed under $+$ i.e. $v_1, v_2 \in V' \Rightarrow v_1 + v_2 \in V'$

(ii) V' is closed under scalar multiplication i.e. $\alpha \in \mathbb{R}, v \in V' \Rightarrow \alpha v \in V'$.

We have already given some eg. of sub-spaces. Here are more:

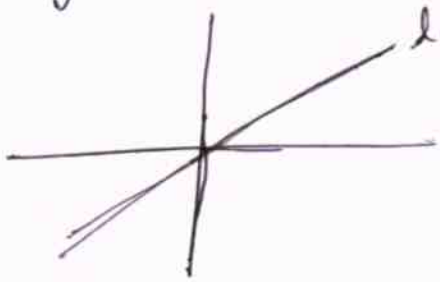
eg i ~~xy~~, vectors in xy or yz or zx planes ~~are~~ form subspace in $\mathbb{R}^3$.

eg ii vectors ~~in~~ on x, y, z axis form a subspace in $\mathbb{R}^3$.

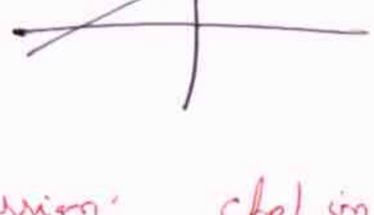
eg iii $\mathbb{V}$ itself is a subspace of $\mathbb{V}$.

eg iv Let $V' = \{0\}$ identity element in $\mathbb{V}$. $(V', +, \cdot)$ is always a subspace. infact it is the "smallest" subspace ($\because$ there can be no vector space without the identity element)

eg v



$l_1 \rightarrow$ any line through origin is a subspace



$l_1 \rightarrow$ any line offset from origin is not a subspace.

Reading assign: chp1 in pdf
sec 2.1 in Gilbert Strang's lin. alg. & appl. book

Lectures 3 and 4

26,27-Oct-09

(Refer Chp2 in pdf for lecturenotes 2)

Till now we were discussing/generalizing the notion of vector spaces and vectors. All said and done, the only operations "permitted" in vector spaces (till now) is "linear combinations". However in case of Euclidean vector spaces, we can talk about length of a vector, distance between vectors and angles between vectors! We will generalize these notions to an abstract vector space by generalizing/defining the concept of an inner product.

As usual, we look at inner product b/w Euclidean vectors and identify ~~[illegible]~~ some fundamental prop. that Euclidean inner product satisfies and preserve them. Later on we will define any function which takes two vectors and gives ~~a~~ number and satisfies the identified fundamental prop. as an inner-product.

The following prop. are fundamental to Euclidean inner products:

$v_3, v_1, v_2 \in V$ where $V = (V, +, \cdot)$ is a vec. sp.

(i) $\langle v_1, v_1 \rangle \geq 0$, $\langle v_1, v_1 \rangle = 0 \Leftrightarrow v_1 = 0$ | non-negativity property

identity element in V

(ii) $\langle v_1, v_2 \rangle = \langle v_2, v_1 \rangle$ | symmetry

(iii) $\langle v_1 + v_2, v_3 \rangle = \langle v_1, v_3 \rangle + \langle v_2, v_3 \rangle$
$\langle \alpha v_1, v_2 \rangle = \alpha \langle v_1, v_2 \rangle$ | linearity in first argument. $\Rightarrow$ linearity in both arguments

Now given an abstract vector space $V = (V, +, \cdot)$ we define a function $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$. If $\langle \cdot, \cdot \rangle$ satisfies all three axioms then $\langle \cdot, \cdot \rangle$ is called an inner product and V is called an inner product space.

eg1 $\mathbb{V} = (\mathbb{R}^n, +, \cdot)$ equipped with usual $\langle \cdot, \cdot \rangle$ Euclidean inner product i.e. $\langle v_1, v_2 \rangle = v_1^T v_2$ is indeed an inner product space!

eg2 $\mathbb{V} = (\mathbb{R}^n, +, \cdot)$ equipped with ~~usua~~ $\langle \cdot, \cdot \rangle_Q$ defined as:

$$\langle v_1, v_2 \rangle_Q = v_1^T Q v_2 \quad \text{where } Q \text{ is a pd matrix.}$$

It is easy to verify that $\langle \cdot, \cdot \rangle_Q$ satisfies all 3 axioms.

However, there is one more "nice" thing happening here.

Q is pd $\iff \exists$ an orthonormal matrix L and diagonal $D \ni Q = LDL^T$
(we know this from Notes of prob. theory)

$\Rightarrow \langle v_1, v_2 \rangle_Q = v_1^T Q v_2 = v_1^T L D L^T v_2 = \langle v_1', v_2' \rangle \longrightarrow$ usual Euclidean inner product !!

where $v_i' = D^{1/2} L^T v_i \longrightarrow$ this transformation of vectors ~~tra~~ is equivalent to rotating & scaling the co-ordinate axis

In other words,

the new inner product between vectors ~~can~~ also represents "usual" inner product in a transformed vector space!

So this gives a way to computed inner products in (lin) transformed inner product spaces! without having to actually do the transformation!

*[Extension of this idea to non-linear transformations is exploited in all ~~kernel~~ Kernel based methods like SVMs!]

Before giving more examples one can realize that an inner product naturally gives a definition for length (norm) of a vector:

Here is what it happens in case of Euclidean vectors:

$\|v\| = \sqrt{v^T v} = \sqrt{\langle v, v \rangle}$.

Encouraged by this can we define $\|v\|$ as $\sqrt{\langle v, v\rangle}$ ~~for~~ in any inner product space?

(first of all we know that $\langle v, v\rangle \geq 0 \ \forall v \in V$, hence is a valid function which takes ~~on~~ a vector and gives a number)

In order to answer this, we must first know what are the fundamental properties satisfied by a norm! Again we look at Euclidean norm, pick fundamental prop.:

For any ~~inner product~~ vector space (need not be inner product space), if we have a function $\|\cdot\| : V \to \mathbb{R}$ such that:

(i) $\|v\| \geq 0 \quad \forall v \in V, \quad \|v\| = 0 \iff v = 0$ / non-negativity

(ii) $\|\alpha v\| = |\alpha| \|v\| \quad \forall \alpha \in \mathbb{R}, v \in V$ / proper-scaling

(iii) $\|v_1 + v_2\| \leq \|v_1\| + \|v_2\|$ / triangle-inequality.

then we calls that vector space a normed vector space and $\|\cdot\|$ as a norm!

Let us now see if the norm function induced by inner product satisfies all these properties or not:

i.e. $\|v\| \equiv \sqrt{\langle v, v\rangle}$

(i) and (ii) are trivially true by the very defn. of $\langle \cdot, \cdot \rangle$. (iii) is to be verified:

TST $\|v_1+v_2\| \le \|v_1\| + \|v_2\|$

i.e. TST $\left(\|v_1+v_2\|\right)^2 \le \left(\|v_1\| + \|v_2\|\right)^2$

i.e. TST $\langle v_1+v_2, v_1+v_2\rangle \le \|v_1\|^2 + \|v_2\|^2 + 2\|v_1\|\|v_2\|$

$$= \langle v_1, v_1\rangle + \langle v_2, v_2\rangle + 2\|v_1\|\|v_2\|$$

i.e. TST $\langle v_1, v_1\rangle + \langle v_1, v_2\rangle + \langle v_2, v_1\rangle + \langle v_2, v_2\rangle = \langle v_1, v_1\rangle + \langle v_2, v_2\rangle + 2\|v_1\|\|v_2\|$

(twice apply linearity prop.)

i.e. TST $\langle v_1, v_2\rangle \le \|v_1\|\|v_2\|$ → (rings a bell? it looks like ~~is the~~ Cauchy-Schwarz inequality)

(symmetry of $\langle\rangle$ and cancel terms)

Proof (again similar to how we did in prob. theory)

we have $\left(\|\alpha v_1 + v_2\|\right)^2 \ge 0 \quad \forall\ \alpha \in \mathbb{R}$ (say v_1, v_2 are given arbitrary vectors)

$\Rightarrow \langle \alpha v_1 + v_2, \alpha v_1 + v_2\rangle \ge 0 \quad \forall\ \alpha \in \mathbb{R}$

$\Rightarrow \alpha^2\langle v_1, v_1\rangle + 2\alpha\langle v_1, v_2\rangle + \langle v_2, v_2\rangle \ge 0 \quad \forall\ \alpha \in \mathbb{R}$

(put $\alpha = 0$, $\langle v_2, v_2\rangle \ge 0$ so it is feasible)

$\Rightarrow$ discriminant $\le 0 \Rightarrow 4\langle v_1, v_2\rangle^2 - 4\langle v_1, v_1\rangle\langle v_2, v_2\rangle \le 0$

$\Rightarrow \langle v_1, v_2\rangle^2 \le \|v_1\|^2\|v_2\|^2$

$\Rightarrow |\langle v_1, v_2\rangle| \le \|v_1\|\|v_2\|$ → Cauchy-Schwarz inequality!

$\Rightarrow \langle v_1, v_2\rangle \le \|v_1\|\|v_2\|$

Hence proved.

So triangular inequality follows from and hence $\|\cdot\|$ is indeed a valid norm!!

Hence any inner product space is a normed vector space equipped with the induced norm $||v|| = \sqrt{\langle v, v\rangle}$.

~~not~~ nothing but generalization of length of vector.

~~Once ordm &<~~

Once $\langle \cdot, \cdot \rangle$ and induced norm $||\cdot||$ are there, concepts of angles and distances between vectors $v_1, v_2 \in V$ follow:

$$\cos\theta = \frac{\langle v_1, v_2 \rangle}{||v_1|| \; ||v_2||}$$

(Note that $|\cos\theta| \le 1$ by Cauchy-Schwartz) hence valid defn.

cosine of angle between vectors v_1, v_2.

$$||v_1 - v_2|| = \sqrt{\langle v_1 - v_2, v_1 - v_2 \rangle}$$

distance between v_1, v_2

So we can talk ~~of~~ about geometrical objects and geometrical problems in any inner-product space! IIIly in a normed vector space we can describe shapes & problems involving ~~no~~ only ~~norms~~ lengths and distances. (We will see examples).

eg1 Consider $V = (\mathbb{R}^n, +, \cdot)$.

we already saw:

$\langle v_1, v_2 \rangle = v_1^T v_2$ (Euclidean inner product)

$|| v_1 || = \sqrt{v_1^T v_1}$ (Euclidean norm) induced norm

$\langle v_1, v_2 \rangle_Q = v_1^T Q v_2$ (another kind of inner product)

$||v_1||_Q = \sqrt{v_1^T Q v_1}$ (induced norm)

Let $v = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

we also know about other norms in $\mathbb{R}^n$: eg:

$||v||_1 = |x_1| + |x_2| + \cdots + |x_n|$ → 1-norm.

In general,

$$||v||_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

→ p-norm.

Verify that $\|\cdot\|_p$ is indeed a valid ~~no~~ norm.

(soon you will realize that proving is not so easy. Try searching for proves, with some tricks it is easy!)

Put $p=1$ gives 1-norm

$p=2$ gives 2-norm $\longrightarrow$ nothing but Euclidean norm

|||ly $\lim_{p\to\infty} \|v\|_p = \|v\|_\infty \longrightarrow$ ~~inf~~ infinity norm

$= \max_{i=1 \text{ to } n} |x_i|$

Also, there are valid norms which can be defined on $\mathbb{R}^n$, but there are not an inner product induced norms! (except $p=2$)

Just to summarize:

$(\mathbb{R}^n, +, \cdot, \langle\,\rangle)$ is usual Euclidean inner product space & induced norm is Euclidean norm.

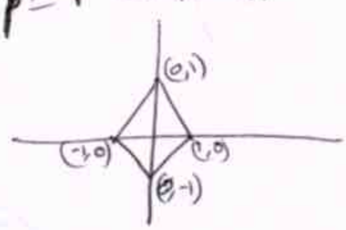
$(\mathbb{R}^n, +, \cdot, \|\cdot\|_p)$ is ~~usual~~ p-norm based Euclidean normed vector space.

$\longrightarrow$ In this space we can only talk abt lengths and distances (no concept of angles!) (except $p=2$)

Just to give more clarity ask the question how will a circle look like when $p=1, 2, \infty$?

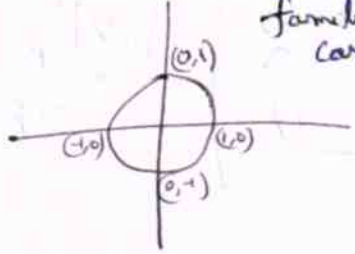
(Recall that circle is the locus of all points with length 1)

$p=1 \Rightarrow \|v\|_1 = 1$

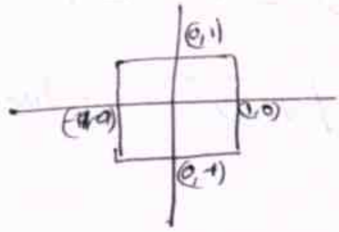


$p=2 \Rightarrow \|v\|_2 = 1$

familiar case!



$p=\infty$ $\|v\|_\infty = 1$



III'ly think about describing the problems of distance between two circles, ellipses, distances between ellipses. So on...
(repeat for every following examples).

<u>eg 2</u> Consider the vector space of all $m \times n$ matrices. After a little thought we will realize that this vector space is equivalent to usual Euclidean vector space $\mathbb{R}^{mn}$! (consider writing elements in matrix as a column vector either row/column major order!)
realize also that $+, \cdot$ defns. "go thru".

This motivates us to define all the inner products, norms we saw in case of Euclidean vectors!

<u>For eg</u> $\langle M_1, M_2 \rangle$ we would define as the Euclidean inner product of those vectors got by unwrapping the matrices M_1 and M_2!

After a little thought you can see that $\langle M_1, M_2 \rangle = \text{trace}(M_1^T M_2)$

(trace of matrix is sum of diagonal terms)

<u>proof in case of 2x2 matrices</u>

suppose $M_1 = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}, M_2 = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix}$

$$v_1 = \begin{bmatrix} x_{11} \\ x_{12} \\ x_{21} \\ x_{22} \end{bmatrix}, \quad v_2 = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} \Rightarrow \langle v_1, v_2 \rangle = x_{11}y_{11} + x_{12}y_{12} + x_{21}y_{21} + x_{22}y_{22}$$

unwrapped vectors (row major order)

$$\text{trace}(M_1^T M_2) = \text{trace}\left(\begin{bmatrix} x_{11}y_{11} + x_{21}y_{21} & \cdots \\ \cdots & x_{12}y_{12} + x_{22}y_{22} \end{bmatrix} \right) =$$

Here $<M_1, M_2> = \text{trace}(M_1^T M_2)$ is analogous to Euclidean inner product & in matrix terminology known as Frobenius inner product!

Also can show $= \text{trace}(M_1 M_2^T)$
hence $= \text{trace}(M_2 M_1^T)$
$= \text{trace}(M_2^T M_1)$

Once Frobenius inner product is defined
Frobenius norm is the induced norm:

$$\|M\| = \sqrt{<M, M>} = \sqrt{\text{trace}(M^T M)}$$

Later on see connection with eigenvalues of $M^T M$.

Since these are "Frobenius" we usually put:
$< \cdot, \cdot >_F$, $\| \cdot \|_F$ (subscripts)

Now one can also extend the definition of p-norm to Matrices also (again by expanding matrices into vectors and then taking usual p-norm)

$$\|M\|_p = \left(\sum_i \sum_j |x_{ij}|^p \right)^{1/p}.$$

One can also talk about angle between matrices & distance between matrices

$$\cos\theta = \frac{<M_1, M_2>_F}{\|M_1\|_F \, \|M_2\|_F} \; ; \; \|M_1 - M_2\|_F = \sqrt{<M_1 - M_2, M_1 - M_2>_F}$$

was used to measure similarity between kernels in SVM!

eg3 Consider the vector space of all r.v.s with finite second moments.
by Cauchy Schwartz inequality $\underbrace{|E\{XY\}|}_{\text{must be } < \infty} \le \underbrace{\sqrt{E\{X^2\}}}_{<\infty} \underbrace{\sqrt{E\{Y^2\}}}_{<\infty}$

Now define $<X, Y> \equiv E\{XY\}$.
It is easy to verify that all 3 axioms of $<.,.>$ are satisfied.

Here it forms an inner-product space. ~~Now we can see~~

induced norm: $\|X\| = \sqrt{\langle X, X\rangle} = \sqrt{E\{X^2\}}$ (= σ std. dev of X if we consider sub-space of r.v.s with mean zero)

(length of r.v)

$$\cos\theta = \frac{\langle X, Y\rangle}{\|X\|\,\|Y\|} = \frac{E\{XY\}}{\sqrt{E\{X^2\}}\,\sqrt{E\{Y^2\}}}$$ (= ρ_{XY} correlation coeff. if we consider sub-space of r.v.s with mean-zero)

(angle between r.vs)

$$\|X-Y\| = \sqrt{\langle X-Y, X-Y\rangle} = \sqrt{E\{(X-Y)^2\}}$$

(distance between two r.vs)

($\sqrt{\ }$ of on average how much do they differ?)

(~~non-intuitive~~)

In all cases, it is easy to see that all meanings are "intuitive"!

Once we have the notion of angle, we ~~too~~ have the notion of orthogonality (between arbitrary vectors):

$$v_1 \perp v_2 \iff \cos\overset{\to 90^\circ}{\theta} = 0 \iff \langle v_1, v_2\rangle = 0.$$

(orthogonal)

So orthogonal vectors are vectors whose inner product is zero.

A set of orthogonal vectors (vectors orthogonal to each other) spanning a vector space is called orthogonal basis.

lly A set of orthonormal vectors (vectors ortho. to each other & length 1) spanning a vec. sp. is called orthonormal basis.

It can be shown that every finite-dimensional vector space has an orthonormal (hence an orthogonal) basis.

Proof is again by construction: idea is to take any basis and constructing the orthonormal basis. (refer Gram Schmidt algorithm)

Of course, again, any set of orthonormal (orthogonal) vectors of 'n' in number span the entire vector space (of dim. n) and hence an orthonormal (orthogonal) basis of it.

Also, since it is a basis, any vector can be uniquely represented in terms of these vectors:

Let $\{e_1, e_2, \ldots e_n\}$ be an orthonormal basis $\Rightarrow \begin{bmatrix} e_i \perp e_j & i \neq j \\ \|e_i\| = 1 & \forall i \end{bmatrix}$

then $V = \alpha_1 e_1 + \alpha_2 e_2 + \cdots + \alpha_n e_n \quad \forall\, v \in V$

(α_i are unique for given v)

Now, $\langle v, e_1 \rangle = \alpha_1$ (since $\langle e_i, e_j \rangle = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$)

Illy. $\langle v, e_i \rangle = \alpha_i \longrightarrow$ so the coeff. are nothing but inner products with the orthonormal basis vectors!

Most important properties of orthonormal basis follow from this observation.

Hilbert space is an inner product space ~~with~~ which is "Complete" (no holes as IIly to $\mathbb{R}^n$)!

Hence Hilbert space is a "full-scale" model of Euclidean space!!

Lecture - 5

29-Oct-09

Till now we looked at ~~vector~~ Euclidean vector spaces and modeled them. We saw that Hilbert space is the generical model for Euclidean vector spaces. Now lets try and see how two vector spaces would interact. To be specific lets consider transformations (or mappings or functions) which take vectors in one space and give vectors in the other:

$$T : V \longrightarrow W \quad \text{where } (V, +, \cdot) \text{ are } (W, +, \cdot)$$

two vector spaces.

"need not be similar (same)" infact does'nt make sense to relate them!

We will immediately give examples:

eg 1 Consider the transformation induced by a matrix $A_{m\times n}$ on the vector space $\mathbb{R}^n$

$$A_{m\times n}\, x_{n\times 1} = y_{m\times 1} \longrightarrow \text{belongs to vector space } \mathbb{R}^m.$$

In other words, the matrix A is taking vectors x in $\mathbb{R}^n$ & giving vectors y in $\mathbb{R}^m$.

eg 2 Consider a fixed vector $v_f \in \mathbb{R}^n$. Now define $T(v) \equiv \langle v, v_f\rangle$ for $v \in \mathbb{R}^n$

It is easy to see that $T : \mathbb{R}^n \longrightarrow \mathbb{R}$ (→ also a vector space!)

eg 3 define $T(v) = \|v\|$ for $v \in \mathbb{R}^n$ (Euclidean norm)

It is easy to see that $T : \mathbb{R}^n \longrightarrow \mathbb{R}$

Now each of the T's defined in the examples are indeed examples of transformations from vectors ~~f~~ in one vector space to the other.

However there is a fundamental difference between eg3 and the other two:

eg1, eg2 are examples of "linear transformations"

i.e. transformations where ~~doing~~ taking the transformation of a linear comb. of vectors is same as linearly combining the transformed vectors. In other words, the order in which the transformation and linear comb. are taken does'nt matter.

i.e. T is a linear transformation iff:

$$\underbrace{T}_{\text{linear transformation}}\big(\underbrace{\alpha . v_1 + \beta . v_2}_{\text{linear combin in } (V,+,\cdot) \text{ space}}\big) = \underbrace{\alpha . T(v_1) + \beta . T(v_2)}_{\text{linear comb. in } (W,+,\cdot) \text{ space!}}$$

$\in V$ $\in W$

It is easy to verify that ~~eg~~ T's in eg1, 2 are indeed linear:

eg1 $A(\alpha v_1 + \beta v_2) = \alpha A v_1 + \beta A v_2$

---

eg2 $< \alpha v_1 + \beta v_2 , v_f > = \alpha < v_1, v_f > + \beta < v_2, v_f >$

Also in case of T in eg 3, this is not true:

$$\| \alpha v_1 + \beta v_2 \| \neq \alpha \| v_1 \| + \beta \| v_2 \|$$

(easy to ~~prove~~ see from triangular inequality $\|x_1 + x_2\| \leq \|x_1\| + \|x_2\|$ $<$ is possible!)

Study of linear transformations between finite dimensional vector spaces is the topic of linear algebla.

(for us it is real vector spaces)

$\longrightarrow$ Lets give more eg. of linear transformations:

<u>eg 4</u> Consider the vector space of all polynomials (over $\mathbb{R}$) of degree less than or equal to m (it is finite dim.) — denote poly. by set P_m (as we prev. noted.)

Consider the $T: P_m \longrightarrow P_{m-1}$ which is nothing but the differential operator.

Let $p_1, p_2 \in P_m$

we know $\dfrac{d}{dx}\left(\alpha p_1(x) + \beta p_2(x)\right) = \alpha \dfrac{d}{dx} p_1(x) + \beta \dfrac{d}{dx} p_2(x)$

(Linearity of transformation)

~~eg5 Consider the vector space of all r.v.s with finite first moments (this is an ∞ dim. vec. space)~~
~~Consider the T: V → ℝ given by the Expectation~~

<u>eg5</u> Consider $T: P_m \longrightarrow \mathbb{R}$ given the "integral in $[a,b]$"

i.e. $T(p_1) = \displaystyle\int_a^b p_1(x)\,dx$.

we know $\displaystyle\int_a^b \left(\alpha p_1(x) + \beta p_2(x)\right) dx = \alpha \int_a^b p_1(x)\,dx + \beta \int_a^b p_2(x)\,dx$

(linearity of integral op.)

<u>eg6</u> Consider $V = (V, +, \cdot)$ the vector space of all r.v.s with finite first moment. (this is an ∞ dim vec. space)

Consider the $T: V \longrightarrow \mathbb{R}$ given by the "Expectation operator" we know $$\underbrace{E[\alpha X+\beta Y] = \alpha E[X]+\beta E[Y]}_{\text{linearity of expectation.}}$$

We can keep giving examples of linear transformations. However linear transformations of between ~~de~~ finite dimensional vector spaces are special and intuitively lead to defining the vector spaces of matrices and various other operators on matrices.

Here is it:

Let $\mathbb{V} = (V, +, \cdot)$ and $\mathbb{W} = (W, +, \cdot)$ be two vector spaces.

(disambiguate based on situation)

Let dim $\mathbb{V} = n$ and dim $\mathbb{W} = m$.

Let $T : V \longrightarrow W$ be a linear transformation.

Now if we have to specify T then we need to specify

$T(v)\ \forall v \in V$ there might be uncountable no. of them
"will" (unless trivial case)

So is there an easy way to specify T?

<u>Yes</u>. Let $\{v_1, v_2, \dots, v_n\}$ be a basis of $\mathbb{V}$

" $\{w_1, w_2, \dots, w_m\}$ " " of $\mathbb{W}$.

Now any $v \in V$ has a unique representation in terms of :

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$$

by linearity of T we have:

$$T(v) = T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n)$$

In other words, if I specify $T(v_1), \ldots, T(v_n)$ i.e. values of T at n points, then T is completely specified!

Moreover, ~~suppose~~ we can specify $T(v_i) \in W$ ~~in terms~~ as linear comb. of $w_1, w_2, \ldots, w_m$

i.e. let

$$\begin{aligned} T(v_1) &= \beta_{11} w_1 + \beta_{21} w_2 + \cdots + \beta_{m1} w_m \\ T(v_2) &= \beta_{12} w_1 + \beta_{22} w_2 + \cdots + \beta_{m2} w_m \\ &\vdots \\ T(v_n) &= \beta_{1n} w_1 + \beta_{2n} w_2 + \cdots + \beta_{mn} w_m . \end{aligned}$$

Convenient to represent as matrix:

remember V, W are abstract vec. spa. However matrix is naturally appearing! motivation for defining matrices!

$$\text{Transformation matrix} = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2n} \\ \vdots & \vdots & & \vdots \\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mn} \end{bmatrix}_{m \times n}$$

$$\begin{aligned} T(v) &= \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n) \\ &= \alpha_1 (\beta_{11} w_1 + \cdots + \beta_{m1} w_m) + \cdots + \alpha_n (\beta_{1n} w_1 + \cdots + \beta_{mn} w_m) \\ &= \underbrace{(\alpha_1 \beta_{11} + \cdots + \alpha_n \beta_{1n})}_{\gamma_1,\ \text{say}} w_1 + \cdots + \underbrace{(\alpha_1 \beta_{m1} + \cdots \alpha_n \beta_{mn})}_{\gamma_m} w_m \end{aligned}$$

(motivation for defn. of matrix vector multiplication!

$$\begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2n} \\ \vdots & \vdots & & \\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_m \end{bmatrix}$$

So given the ~~mat~~ transformation matrix, the linearity transformation is specified and vice-versa.

Hence set of all possible transformations from $V \to W$ is nothing but all possible matrices of size $m \times n$!

rings a bell that this will form vec. spa. with suitable +, ·.

Hence motivates to define $T_1 + T_2$ and αT_1 & in general lin. comb. of lin. transformations

i.e. motivates defining addition and scaling of matrices!

So motivated by this we can define

$$(T_1 + T_2)(v) = T_1(v) + T_2(v) \quad \text{nothing but} \quad M_1 + M_2$$

$$(\alpha . T_1)(v) = \alpha . T_1(v) \iff \alpha M_1$$

addition of corresponding transformation matrices

~~expansion~~ & scaling of transformation matrix

eg (i) what is the transformation ~~of~~ matrix of the transformation represented by $A_{m \times n} x_{n \times 1} = y_{m \times 1}$ w.r.t. to the std. basis of $\mathbb{R}^m$ & $\mathbb{R}^n$?

Answer is A itself:

$$\text{std. basis of } \mathbb{R}^m \text{ is } \underset{w_1}{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{m \times 1}}, \underset{w_2}{\begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{m \times 1}}, \ldots, \underset{w_m}{\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}_{m \times 1}} \to m \text{ such vectors}$$

$$\text{'' '' } \mathbb{R}^n \text{ is } \underset{v_1}{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{n \times 1}}, \underset{v_2}{\begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{n \times 1}}, \ldots, \underset{v_n}{\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}_{n \times 1}} \to n \text{ such vectors.}$$

Let the column vectors in matrix A be $a_1, a_2, \ldots, a_n$ (each will be $m\times 1$ vector).

$$T(v_1) = Av_1 = \begin{bmatrix} \vdots & \vdots & & \vdots \\ a_1 & a_2 & \cdots & a_n \\ \vdots & \vdots & & \vdots \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = a_1 = \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1m} \end{bmatrix} = a_{11}\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + a_{1m}\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

$$= a_{11}w_1 + \cdots + a_{1m}w_m$$

nothing but first column of A!

$\Rightarrow$ the transformation matrix is A itself.

Eg (ii) Consider the vec. spa. of all polynomials of degree upto m, and the differential operator as the linear transformation, we know:

$$T : P_m \rightarrow P_{m-1}$$

Now consider the std. basis of P_m : $\{\underset{v_1}{1}, \underset{v_2}{x}, x^2, \ldots, \underset{v_{m+1}}{x^m}\}$

" P_{m-1} : $\{\underset{w_1}{1}, \underset{w_2}{x}, \ldots, \underset{w_m}{x^{m-1}}\}$

$$T(v_1) = \frac{d}{dx}1 = 0 = 0.w_1 + 0.w_2 + \cdots + 0.w_m$$

$$T(v_2) = \frac{d}{dx}x = 1 = 1.w_1 + 0.w_2 + \cdots + 0.w_m$$

$$T(v_3) = \frac{dx^2}{dx} = 2x = 0.w_1 + 2.w_2 + \cdots + 0.w_m$$

$$\vdots$$

$$T(v_{m+1}) = \frac{dx^m}{dx} = mx^{m-1} = 0.w_1 + 2.w_2 + \cdots + m.w_m$$

$\Rightarrow$ Transformation matrix is $\begin{bmatrix} 0 & 1 & 0 & & 0 \\ 0 & 0 & 2 & & \vdots \\ \vdots & 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & & 0 \\ 0 & 0 & 0 & & m \end{bmatrix}$

Irrespective of what vector spaces are involved in a linear transformation between finite dim. vec. spa., the lin. tra. can be represented by a matrix (of suitable dim.) and set of all such matrices give all possible lin. tra.

So studying linear trans. f boils down to study of <u>Matrices</u> and operations on them!

As mentioned earlier various operations of matrices can be motivated from their meanings/implications wrt. the lin. tra. (we already saw $+, \cdot$)

Composition of two Transformations leads to the concept of matrix multiplication!

$$\begin{array}{ccc} T_1 : \underset{n\,\text{dim}}{V} \to \underset{m\,\text{dim}}{W} & & T_2 : \underset{m\,\text{dim.}}{W} \longrightarrow \underset{p\,\text{dim}}{U} \\ \downarrow & & \downarrow \\ M_{T_1} \longrightarrow \text{size } m \times n & & M_{T_2} \longrightarrow \text{size } p \times m \\ \left(\begin{array}{c}\text{trans. matrix} \\ \text{of } T_1\end{array}\right) & & \left(\begin{array}{c}\text{tran. matrix} \\ \text{of } T_2\end{array}\right) \end{array}$$

Consider $T_2 \circ T_1 : V \to U$

Tra. matrix of just this is given by $M_{T_2} M_{T_1}$! → usual matrix multiplication (check this!)

of course, unless $n = p$, $T_1 \circ T_2$ is not even defined
(and no is $M_{T_1} M_{T_2}$!)

In fact we can also motivate definition of matrix inverse by looking at invertibility of a lin. transformation! In order to do that let us study more about matrices:

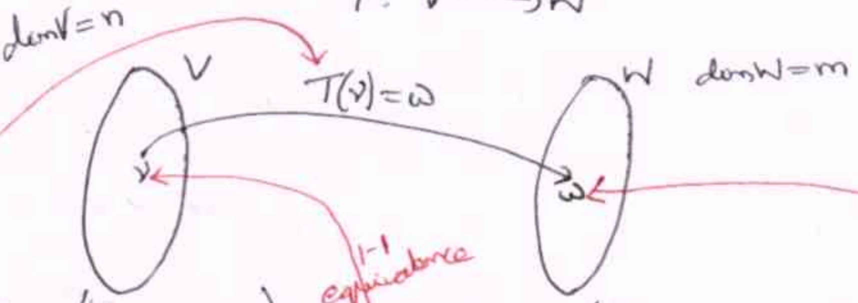
Lecture Notes - 6

1-Nov-09

We discussed in last lecture that specifying a lin. tra. (and studying it) is equivalent to specifying a tra. matrix (of appropriate size and studying it). Here is a pictorial view of the same:

$\mathbb{V}=(V,+,\cdot)$ $\qquad\qquad$ $\mathbb{W}=(W,+,\cdot)$

$$T: V \longrightarrow W$$



$\mathbb{V}$ & $\mathbb{W}$ might be abstract vector spaces!

1-1 equivalence

1-1 equivalence

1-1 equivalence

$V = \text{span}(\underbrace{\{v_1, v_2, \ldots, v_n\}}_{\text{basis}})$ $\qquad$ $W = \text{span}(\underbrace{\{\omega_1, \omega_2 \ldots \omega_m\}}_{\text{basis}})$

Specifying $T(v)\ \forall v$ is equivalent to specifying $T(v_1), \ldots T(v_n)$ (we saw in prev. lecture)
The transformation matrix helps to specify in terms of basis of $\mathbb{W}$:

each column tells how to combine basis of $\mathbb{W}$ to get corresponding $T(v_i)$

$$\underbrace{\begin{bmatrix} \beta_{11} & \beta_{12} & & \beta_{1n} \\ \beta_{21} & \beta_{22} & & \beta_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \beta_{m1} & \beta_{m2} & & \beta_{mn} \end{bmatrix}}_{\text{say, } M} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}}_{\text{say, } \bar{\alpha}} = \underbrace{\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_m \end{bmatrix}}_{\text{say, } \bar{\gamma}}$$

M ↓ transformation matrix

$\bar{\alpha}$ ↓ specifies how to combine basis of $\mathbb{V}$ to get v

$\bar{\gamma}$ → specifies how to combine basis of $\mathbb{W}$ to get ω.

So, studying lin. tra. of finite dim (abstract) vec. spa. is equivalent to studying tra. mat. and operations/properties of it!

T is one-one iff $\begin{matrix} T(v_i) = w \\ T(v_j) = w \end{matrix} \Rightarrow v_i = v_j$

$$\iff \begin{matrix} M\bar{\alpha}_1 = \bar{y} \\ M\bar{\alpha}_2 = \bar{y} \end{matrix} \Rightarrow \bar{\alpha}_1 = \bar{\alpha}_2$$

$$\iff M(\bar{\alpha}_1 - \bar{\alpha}_2) = 0 \Rightarrow \bar{\alpha}_1 - \bar{\alpha}_2 = 0$$

$\iff$ all columns of M are LI.

T is onto iff $\exists$ a $v \in V \ni T(v) = w \quad \forall w \in W$

$$\iff \exists \text{ a } \bar{\alpha} \ni M\bar{\alpha} = \bar{y} \quad \forall \bar{y} \in \mathbb{R}^m$$

$\iff$ columns of M span $\mathbb{R}^m$

Let just take some cases of specific kinds of M:

(i) $\underline{m > n}$ (no. of columns of M is n) so this is the case where span of columns of M can at most be $\mathbb{R}^n$!

$\therefore$ colums of M cannot span $\mathbb{R}^m$

$\therefore$ In this case any M <u>cannot</u> represent an onto transformation

However, M may or may not represent an 1-1 transf. based on LI of columns of M.

(ii) $n > m$ (this is the case where the columns of M can never be LI (because there are more columns than m itself)

$\Rightarrow M$ <u>cannot</u> represent a 1-1 tra.

However, M may or may not represent an onto tra. based on whether cols of M span $\mathbb{R}^m$ or not!

(iii) $n = m$, here cols of M LI $\iff$ cols. of M span $\mathbb{R}^m$

$\therefore$ either it is a tra. which is bijective or not 1-1, not onto!

In case $n = m$ & all cols. of M $\overset{\text{mod}}{\angle} I$, then that tra. is bijective and the tra. matrix of the corresponding inverse tra. is defined to be the <u>inverse of matrix</u> $M \to$ original tra. mat.

↳ check that this matches with school defn. of inverse atleast for $n=m=2,3$.

Now motivated by the use of the behaviours of cols. of M we can define the following:

ⓐ <u>Column Space of M</u> : span of cols. of M (notation $\mathcal{C}(M)$)

by defn. it is a vector space. However it is subset of $\mathbb{R}^m$.

$$\therefore \; \underset{\text{(subspace)}}{\mathcal{C}(M) \subseteq \mathbb{R}^m}$$

ⓑ Nullspace of T : All vectors in V which map to identity element in W

or $\Updownarrow$

<u>Nullspace of M</u> : All vectors $\bar{x} \ni M\bar{x} = 0$. (notation $\mathcal{N}(M)$)

$$\mathcal{N}(M) \subseteq \mathbb{R}^n \qquad \mathcal{N}(M) = \{x \in \mathbb{R}^n / Mx = 0\}$$

Moreover, $\mathcal{N}(M)$ is closed under add & sca. mult. so it is infact subspace of $\mathbb{R}^n$!

$$\underset{\text{(subspace)}}{\mathcal{N}(M) \subseteq \mathbb{R}^n}$$

ⓒ Row space of M or <u>Col. space of M^T</u> : span of rows of M (or) span of cols. of M^T. (notation $\mathcal{R}(M)$)

by defn. it is vectorspace & subset of $\mathbb{R}^n$.

$$\therefore \underset{\text{(subspace)}}{\mathcal{R}(M) \subseteq \mathbb{R}^n}$$

(d) Null space of M^T or Left null space of M : all vectors $\bar{x} \ni M^T \bar{x} = 0$ (denoted $\mathcal{N}(M^T)$)

$\mathcal{N}(M^T) = \{x \in \mathbb{R}^m / M^T x = 0\}$

Again $\longrightarrow$ $\mathcal{N}(M^T) \subseteq \mathbb{R}^m$ (subspace)

These four subspaces are known as the fundamental subspaces of a matrix.

It is easy to see that $\mathcal{N}(M) \perp \mathcal{R}(M) = \mathcal{C}(M^T)$

(read as null space is ~~ago~~ orthogonal to row space

||ly $\mathcal{N}(M^T) \perp \mathcal{R}(M^T) = \mathcal{C}(M)$

$\longrightarrow$ we know that $v_1 \perp v_2 \iff \langle v_1, v_2 \rangle = 0$ (orthogonality of vectors)

Orthogonality of any two sets of vectors is also defined analogously:

$V_1 \perp V_2 \iff \forall v_1 \in V_1, \forall v_2 \in V_2$ we have $v_1 \perp v_2$.

$\longrightarrow$ Now lets prove $\mathcal{N}(M) \perp \mathcal{R}(M)$:

~~by defn~~. Let x be in $\mathcal{N}(M)$ then by defn. $Mx = 0$ $\to$ vector zero

$$\text{rows of } M \Rightarrow \begin{bmatrix} M_1^T \\ M_2^T \\ \vdots \\ M_n^T \end{bmatrix}_{m \times n} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{n \times 1} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{m \times 1}$$

M_i^T is i^{th} row vector of M

$\Rightarrow M_i^T x = 0 \;\; \forall i \Rightarrow x$ is $\perp$ to all rows of M

$\to$ number zero

$\Rightarrow x$ is $\perp$ span of rows of M

Since x is arbitrary, we have $\mathcal{N}(M) \perp \mathcal{R}(M)$ $\Rightarrow x \perp \mathcal{R}(M)$

In fact, one can prove ~~…~~:

(i) $\mathcal{N}(M) \perp \mathcal{R}(M)$ (already proved!)

(ii) $\underbrace{\dim \mathcal{N}(M)}_{\text{subspace of } \mathbb{R}^n} + \underbrace{\dim \mathcal{R}(M)}_{\text{subspace of } \mathbb{R}^n} = n$

} Fundamental Theorem of Linear Algebra.

(or fund. theo. of matrices)

If for any two subspaces (i) & (ii) hold, they are known as "orthogonal complements" of each other!.

We need to prove (ii) which is not easy to do unless we prove the following:

called RANK-NULLITY THEOREM

$\underbrace{\dim \mathcal{N}(M)}_{\text{subspace of } \mathbb{R}^n} + \underbrace{\dim \mathcal{C}(M)}_{\text{subspace of } \mathbb{R}^m} = n$ → this is easy to prove! Later we get (ii) from this statement.

→ note the diff between (ii) & this!

Proof: $\mathcal{N}(M)$ is a subspace so it will have a basis. Let $\dim \mathcal{N}(M)$ be d

$\Rightarrow$ Let $\mathcal{N}(M) = \text{span}(\underbrace{\{n_1, n_2, \dots n_d\}}_{\text{basis}})$

Now we can extend this basis by adding $n-d$ vectors LI to these & each other such that they span entire $\mathbb{R}^n$.

$\Rightarrow$ Let $\mathbb{R}^n = \text{span}(\underbrace{\{n_1, \dots, n_d, u_1, u_2, \dots u_{n-d}\}}_{\text{basis of } \mathbb{R}^n})$

Now any $x \in \mathbb{R}^n$ can be uniquely written as lin. comb. of :

Since x is arbitrary, Mx represents an arbitrary vector in $\mathcal{C}(M)$

but $Mx = \cancel{M(\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_d u_d + \alpha_{d+1} n_1 \cdots)}$

$$M(\alpha_1 n_1 + \alpha_2 n_2 + \cdots + \alpha_d n_d + \alpha_{d+1} u_1 + \alpha_{d+2} u_2 + \cdots + \alpha_n u_{n-d})$$

$$\Rightarrow Mx = \underbrace{\alpha_1 Mn_1 + \alpha_2 Mn_2 + \cdots + \alpha_d Mn_d}_{0 \text{ by defn. of null-space}} + \alpha_{d+1} Mu_1 + \cdots + \alpha_n Mu_{n-d}$$

$$\Rightarrow Mx = \alpha_{d+1} Mu_1 + \cdots + \alpha_n Mu_{n-d}$$

$\Rightarrow Mu_1, \ldots Mu_{n-d}$ span the subspace $\mathcal{C}(M)$

Now if we prove $Mu_1, \ldots Mu_{n-d}$ are LI then we would have they are basis of $\mathcal{C}(M) \Rightarrow \dim \mathcal{C}(M) = n-d$, Hence the Rank-nullity theorem will be proved.

It is easy to prove $Mu_1, \ldots, Mu_{n-d}$ are LI:

Proof suppose they are LD, then $\beta_1 Mu_1 + \cdots + \beta_{n-d} Mu_{n-d} = 0$ where not all β_i are zero.

$$\Rightarrow M(\beta_1 u_1 + \cdots + \beta_{n-d} u_{n-d}) = 0$$

$$\Rightarrow \beta_1 u_1 + \cdots + \beta_{n-d} u_{n-d} \text{ is in } \mathcal{N}(M)$$

which is a contradiction to very defn. of $u_1, \ldots, u_{n-d}$.

Hence Rank-nullity theorem is proved.

So we have:

$$\dim \mathcal{N}(M) + \dim \mathcal{C}(M) = n \quad \text{III}$$

$\mathcal{N}(M) \subseteq \mathbb{R}^n$, $\mathcal{C}(M) \subseteq \mathbb{R}^m$

$\dim \mathcal{C}(M)$ called column rank (r_c)

i.e. $\dim \mathcal{C}(M) = r_c$

Applying this theorem to the matrix M^T we have:

$$\dim \mathcal{N}(M^T) + \dim \mathcal{C}(M^T) = m \quad \text{IV}$$

$\mathcal{N}(M^T) \subseteq \mathbb{R}^m$, $\mathcal{C}(M^T) \subseteq \mathbb{R}^n$

$\dim \mathcal{C}(M^T)$ called row rank (r_r)

i.e. $\dim \mathcal{C}(M^T) = \dim \mathcal{R}(M) = r_r$

Now since $\mathcal{N}(M)$ and $\mathcal{R}(M)$ are subspaces in $\mathbb{R}^n$ we have:

$$\dim \mathcal{N}(M) + r_r \leq n = \dim \mathcal{N}(M) + r_c \quad \text{by III}$$

$$\Rightarrow r_r \leq r_c$$

Illy $\mathcal{N}(M^T)$ and $\mathcal{C}(M)$ are subspaces in $\mathbb{R}^m$, so we have:

$$\dim \mathcal{N}(M^T) + r_c \leq m = \dim \mathcal{N}(M^T) + r_r \rightarrow \text{by IV}$$

$$\Rightarrow r_c \leq r_r$$

$$\Rightarrow r_r = r_c \longrightarrow \text{this common value is called the } \underline{\text{rank of the matrix}}.$$

i.e. no. LI rows = no. LI cols = r (rank)

Since $\dim \mathcal{C}(M) = \dim \mathcal{R}(M) = r$, we have the fundamental theorem of Linear Algebra proved. (using the rank-nullity theorem)

Lets look at some eg:

<u>eg</u>! Consider the differential operator on vector space of all polynomials upto degree m

we ~~kn~~ know its a lin. tra. ~~[illegible]~~

It is easy to see that this is not 1-1 (all constant polynomials go to zero polynomial)

but it is onto.

This can be realized from the tra. matrix also:

$$M = \begin{bmatrix} 0 & 1 & 0 & & 0 \\ 0 & 0 & 2 & & \vdots \\ \vdots & \vdots & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & & 0 \\ 0 & 0 & 0 & & m \end{bmatrix}_{m \times m+1}$$ (we know this from prev. lecture)

because of the zero column, the columns are LD

$\Rightarrow$ not 1-1 ✓

However rest of columns are LI & they are m of them $\Rightarrow$ rank = m

$\Rightarrow$ span $\mathbb{R}^m$

$\Rightarrow$ <u>onto transformation</u> ✓

(dim of P_{m-1} is indeed m)

set of all poly. upto degree m-1

Also now we can talk about

$\dim N(M) = (m+1) - \underbrace{m}_{\text{rank}} = 1$

Also $\dim N(M^T) = m - m = 0$ $\Longrightarrow$ the adjoint transformation <u>is 1-1</u>

$\hookrightarrow$ defined as the tra. corresponding to M^T

However adj. will not be onto!

as rank $\leq$ <u>m+1</u>.

Since we defined adjoint tra. we can also define the case with symmetric matrix M as self-adjoint transformation.

Also, only in the case where M is orthonormal, we have $M^{-1}=M^T$

i.e. the adj. transformation is the inverse transformation.

$\longrightarrow$

A matrix is invertible iff. $m=n=r$ (the rank)

Also it is easy to see $r \le \min(m,n)$.

we can also look at cases $r=\min(m,n)$ and $m \ne n$

- i.e. $r=m$ (full row rank) $\longrightarrow$ only if $n>m$, i.e. not a 1-1 transformation but onto
- or $r=n$ (full column rank) $\longrightarrow$ only if $m>n$, i.e. not onto but surely 1-1

we can also ask how best we can approximate in case tra. is not onto?

i.e. find $x \in \mathbb{R}^n \ni \|Mx-b\|$ is minimized.

if b is in $\mathcal{C}(M)$ then ↓ is zero. else there will be error.

$$\min_x \|Mx-b\|^2$$

$$\Rightarrow \min_x (Mx-b)^T(Mx-b)$$

$$\Rightarrow \min_x (x^TM^T-b^T)(Mx-b)$$

$$\Rightarrow \min_x x^TM^TMx - 2x^TM^Tb + b^Tb$$

$$\Rightarrow \nabla(\;\;) = 0 \Rightarrow 2M^TMx - 2M^Tb = 0 \Rightarrow \underline{M^TMx = M^Tb}$$

note the optimal solves this linear equation!

This equation $M^TMx = M^Tb$ is called the "normal eqns." corre-
to the original eqn. $Mx = b$

It turns out the normal eqn. always has a solution!!

(irrespective of whether or not $Mx=b$ has a soln.)

It is easy to prove this:

RHS is nothing but a vector in $\mathbb{R}^n$ which is in range of M^T!
LHS is also a vector in range of M^T! So we must be able to find x.

Of course if M has full column rank (i.e. rank $= n$) then $x \to Mx$ is 1-1 (because dim $N(M)=0$)
So the solution of $\underbrace{M^TMx = M^Tb}$ will be unique ($\forall b$)

$\Rightarrow M^TM$ is invertible!

$\therefore$ If $r = n$ (full colm. rank) then M^TM is invertible

& $x = \underbrace{(M^TM)^{-1}M^T}_{\text{called left inverse of } M} b$ gives the required optimal $\boxed{\text{least squares soln.}}$

because $\left((M^TM)^{-1}M^T\right)M = I$

Also no two $x_1 \neq x_2$ will have $M^TMx_1 = M^TMx_2$ because if this happens then $M^T(Mx_1 - Mx_2) = 0$
$\Rightarrow Mx_1 - Mx_2$ is in $N(M^T)$! which is not possible since $Mx_1, Mx_2 \in C(M)$ & are distinct.

Ill'ly if $r = m$ (full row rank), then MM^T is invertible & we can define $M^T(MM^T)^{-1}$ as the right inverse of M

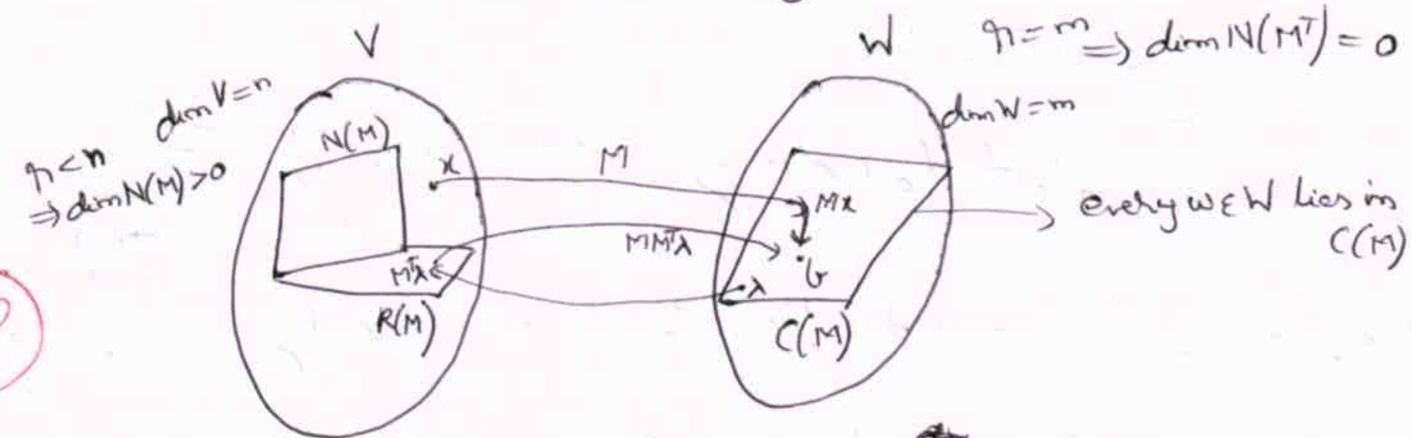
Since $M\left(M^T(MM^T)^{-1}\right) = I$.

Now if $r < \min(m,n)$ can we define a "suitable" inverse? The answer is Yes and this is one motivation for studying Eigen(Singular ~~value~~) value decomposition

Lets work out the $r=m$ (full row rank, $r<n$ rectangular) case also: $Mx=b$

(Convince yourself that case $r=n$ (full column rank, $r<m$ rectangular) represents solving set of eqns. which are more than the unknowns themselves! & of course feasibility is in question here.)

Now in this case [there are more unknowns than equations, "so if the eqns are consistent" which is true if $r=m$, then there will be multiple solns.], we have ~~that~~ the following picture:



Easy to see that $Mx \in C(M)$ & since $N(M^T)=\{0\}$, we have all $w\in W$ are also in $C(M)$ so ~~the~~ $Mx=b$ is always solvable in case of $r=m, r<n$. However there ~~might~~ will be many vectors which fall on b! ~~i.e.~~ this is because $N(M)\neq\{0\}$.

Now which one of these will we select? Ideally I will want to choose the vector which solves $Mx=b$ & is "simple". "Simple" means vector is as close to zero vector as possible. (intuitively zero is the simplest vector). In other words we would like x to be at minimum distance from 0 (i.e. has minimum length) while satisfying $Mx=b$.

i.e. (I) $\boxed{\begin{array}{l} \min \frac{1}{2}\|x\|^2 \\ \underline{\text{s.t.}} \quad Mx = b \end{array}}$ $\longrightarrow$ minimize squared length of x

$\longrightarrow$ while satisfying the constraints that $Mx = b$

Students who know Lagrange multipliers would know

$$\mathcal{L} = \frac{1}{2}\|x\|^2 + \lambda^T(Mx - b)$$

Lagrangian $\qquad \lambda_{m\times 1} \to m$ dim vector of Lagrange multipliers

$$\nabla_x \mathcal{L} = 0 \Rightarrow 2x + M^T\lambda = 0 \Rightarrow \boxed{x = -M^T\lambda} \to \text{eqn. relating unknowns } x \text{ \& Lagrange multipliers}$$

$$\mathcal{L} = \frac{1}{2}(-\lambda^T M)(-M^T\lambda) + \lambda^T M(-M^T\lambda) - \lambda^T b$$

$$= -\frac{1}{2}\lambda^T M M^T \lambda - \lambda^T b$$

(I) is equivalent to solving $\max_{\lambda} \mathcal{L}$ i.e. $\boxed{\min_{\lambda} \frac{1}{2}\lambda^T M M^T \lambda + \lambda^T b}$ (II)

Now once (II) is solved, the optimal x is given by $-M^T\lambda$.

(II) is easy to solve : put grad wrt. $\lambda = 0 \Rightarrow \boxed{MM^T\lambda = -b}$

Now one can show that $MM^T\lambda = -b$ always has a unique solution (in our case i.e. $r = m < n$). Existance is again evident by picture (P)

Uniqueness again follows from the argument given in case of $r = n < m$ which is as follows:

Suppose $\lambda_1 \neq \lambda_2$ are such that $MM^T\lambda_1 = MM^T\lambda_2$

Since $\Rightarrow M(M^T\lambda_1 - M^T\lambda_2) = 0, \quad \lambda_1 \neq \lambda_2$

$\Rightarrow M^T\lambda_1 - M^T\lambda_2 \in N(M), \quad \lambda_1 \neq \lambda_2$

this is impossible because $\in R(M) \perp N(M)$ and $\therefore M^T\lambda_1 \neq M^T\lambda_2$ as $N(M^T) = 0$!

Here in case $r = m < n$, we have that $MM^T\lambda = -b$ is solvable for any b & that too uniquely $\Rightarrow MM^T$ is invertible

$$\lambda = -(MM^T)^{-1} b$$

$$\Rightarrow \boxed{x = -M^T\lambda = M^T\underbrace{(MM^T)^{-1}}\, b}$$

minimum norm solution.

Again, $M\left(M^T(MM^T)^{-1}\right) = I$

However $M^T(MM^T)^{-1}M$ may not be I. So this matrix $M^T(MM^T)^{-1}$ is called the right inverse of M.

<u>In summary:</u>

For $M_{m\times n}$ we have:

(i) $m = n = r$, (full rank) $\Longleftrightarrow$ M is invertible i.e. M^{-1} exists

(ii) $m = r < n$ (full row rank) $\Longleftrightarrow$ (MM^T) is invertible & right inverse $M^T(MM^T)^{-1}$ exists.

(iii) $n = r < m$ (full col. rank) $\Longleftrightarrow$ (M^TM) is invertible & left inverse $(M^TM)^{-1}M^T$ exists.

In general, if I represent a <u>generic matrix inverse</u> as $M^\dagger$

then $x = M^\dagger b$ $\longrightarrow$ is the solution if M is invertible

$\longrightarrow$ is the least square soln. if $n = r < m$

$\longrightarrow$ is the min. norm soln. if $m = r < n$.

Now if $r < m$ & $r < n$, then we will surely not be able to find solns. for all $Mx = b$ and moreover, if we find a least square soln., then the soln. will not be unique!

So what we look for is a minimum norm, ~~min~~ least square soln.

~~One~~ So we would ~~as~~ cast it as an optimization problem with two objectives: (i) min. norm of x (ii) ~~have~~ min. squared error i.e. $\|Ax-b\|^2$

One way to express such problems with two objectives is through the ~~Tikhonov regularization~~ Tikhonov regularization which is simply ~~add~~ a weighted addition of the two objectives:

$$\min_x \; \underbrace{\delta}_{\substack{\text{weighting}\\(\delta>0)}} \underbrace{\|x\|^2}_{\substack{\text{min. norm}\\\text{part}}} + \underbrace{\|Mx-b\|^2}_{\text{least squares part}} \qquad \text{III}$$

High values of δ (say ∞) will give no weight to [least squares part] so will get a soln. that only minimizes $\|x\|$ & we will get ~~zero~~ $x=0$ as answer

Low values of δ (say 0) will give no weight to min. of norm ~~to~~ and give one of the least square ~~minima~~ minimizers.

What we want is hence a solution with $\delta \downarrow 0$, so that we get a unique least square minimizer.

Now III is solved at $\nabla_x \text{ of obj} = 0 \Rightarrow 2\delta x + 2M^TMx - 2M^Tb = 0$

$$\Rightarrow (M^TM + \delta I)x = M^Tb$$

~~&~~ Now it is easy to see that M^TM is a symmetric psd matrix. for any M!

$\because M^TM$ is psd so $\exists$ a L & $D \geq 0$ $\Rightarrow$ $M^TM = LDL^T$.

orthonormal matrix, diagonal matrix

recall that entries on D are eigen-values!

$$\Rightarrow \left(M^TM + \delta I\right) = \left(LDL^T + \delta I\right) = L\underbrace{\left(D + \delta I\right)}_{>0 \text{ since } D \geq 0 \text{ \& } \delta > 0}L^T \quad \text{because } LL^T = I$$

$\Rightarrow M^TM + \delta I$ is pd $\forall \delta > 0$

$\Rightarrow M^TM + \delta I$ is invertible $\forall \delta > 0$

$\therefore$ $x = \left(M^TM + \delta I\right)^{-1} M^T b$ is the min. soln. least square soln.

& as motivated earlier we need the case $\delta \downarrow 0$

$$\lim_{\delta \downarrow 0} \left(M^TM + \delta I\right)^{-1} M^T \longrightarrow \text{defined as } M^{\dagger} \text{ (the generic inverse of a matrix)}$$

As seen even here, the defn. of $M^{\dagger}$ involves eigenvalues. So lets study them.

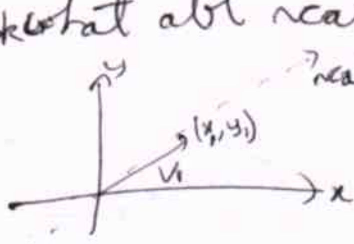
Eigen Value Decomposition

Suppose we ~~ca~~ are considering $T: \mathbb{R}^n \longrightarrow \mathbb{R}^n$ which are linear transformations. Now once we fix a basis for $\mathbb{R}^n$, say the standard basis, then lin. tra.$_n$ is uniquely represented as a square matrix of size n i.e. $M_{n \times n}$.
If the matrix is diagonal, then computing many matrix operations including taking powers & inverses is very easy. In other words diagonal matrices are simple to handle & sometimes non-trivial. Lets look at what transformations do symmetric matrices represent?

eg! $n=2$ $\begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} s_1 x \\ s_2 y \end{bmatrix}$

some 2×2 ~~symmetric~~ diagonal matrix — takes a vector in $\mathbb{R}^2$ — scales it in both directions!

So diagonal matrices represent simple transformations that scale the vectors in the fundamental directions. Once this geometric picture is clear, we can also ask what abt scaling along a particular direction $v_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ i.e.



by say α times

This transformation surely will not be having a ~~symmetric~~ diagonal matrix as the tra. mat. However intuitively this tra. is as simple as in eg! . Infact immediately we will conclude that if the basis of $\mathbb{R}^2$ is chosen as ~~the direct~~ v_1 & v_2 where $v_2 \perp v_1$, then, the tra. mat. will again be diagonal.

So this raises ~~an~~ important questions:

(i) Are all matrices "equivalent" to ~~sy~~ diagonal matrices under suitable basis of $\mathbb{R}^n$? Answer is No (only some matrices are "diagonalizable")

if no

(ii) Can you characterize matrices which are "diagonalizable" i.e. matrices which look "diagonal with suitable basis of $\mathbb{R}^n$.

Ans Yes. (we will indicate later)

In order to answer these questions we will have to look at a fundamental property being satisfied by the basis vectors ~~of~~ ^with a^ diagonal matrix:

Observe that in eg! ~~the x-axis~~ ^vectors lie^ the vectors along x-axis & y-axis are the only vectors which doe not change their orientation after the stretching transformation. Illy in other eg. of scaling/ ~~stretching~~ along v_1, the vectors in direction of v_1 & v_2 are not changing in orientation!! So in order to answer the important questions raised we would like to see if there are directions, i.e. vectors $x \ni \underbrace{Mx}_{\text{vector after lin. tra.}} = \underbrace{\lambda x}_{\text{vector with same orientation as } x.}$

Of course $x=0$ is a trivial soln. $x=0$ doesn't get transformed by any lin. tra. So we look for non-trivial solutions:

$$Mx = \lambda x, \quad x \neq 0$$

If there exists an x satisfies ↑ then λ is called ~~the~~ an eigenvalue of the matrix M

term means "characterizing value" "characteristic value".

Obviously questions about existence of λ, uniqueness of λ etc. etc. pop-up. which we will discuss in next lecture

Lecture Notes 8, 9, 10, 11, 12

We gave many motivations for studying the following problem.

$$\boxed{Mx = \lambda x, \ x \neq 0}$$

If $\exists\ x, \lambda \ni$ this eqn. is satisfied for a given square matrix M, then λ is called an eigenvalue & x is called an eigenvector (corresponding to the eigenvalue λ). The problem itself is called as the eigen-value problem.

Irrespective of the motivations to study this problem, the ~~problem~~ e.v.p. itself is interesting to study:

It gives us orientations invariant under the transformation or in other words, ~~direction~~ vectors for which the lin. tra. represented by M simply looks like a 'scaling' operation.

Before going more in e.v. analysis, ~~to~~ lets right away give examples of evp's ~~appearing~~ in various contexts (applications):

eg1) Let $M = P^T$ where P is the state transition prob. matrix of a finite state Markov chain!

We know for such a chain, stationary distribution(s) exist.

Let Π be a st. dist. $\Pi^T = \Pi^T P$ i.e. $\boxed{P^T \Pi = \Pi}$

This says that Π is nothing but an eigenvector of the matrix P^T corresponding to eigenvalue 1.

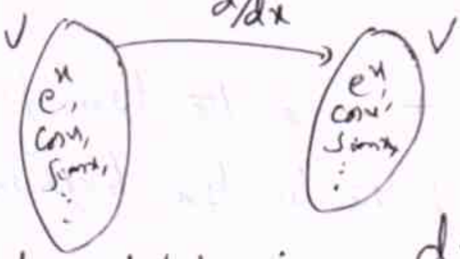
Google solves this particular eigenvalue problem and ranks the results pertaining to a given query! Please look up "PageRank" algorithm.

$\mathbb{V} = (V, +, \cdot)$

eg2 Consider the vector space of all analytic functions. Analytic functions are "natural extensions" of polynomial functions. They are infinitely differentiable & Taylor series expansion around any pt. converges.

eg. of analytic functions are e^x, $\cos x$, $\sin x$ ... so on.

Now consider the differential operator which takes each analytic function & gives another analytic function. We already know it is a linear trans.



infinitely diff. functions are analogous to r.v's with all moments existing.

∴ The e.v.p. relevant here is: $\dfrac{df}{dx} = \lambda f$. If $\exists\, f \neq 0$ satisfies this then the evp has a solution.

analytic fun. are analogous to r.v's with mgf existing!

$\downarrow$ nothing but a 1st order differential eqn.

$$\Rightarrow \frac{df}{f} = \lambda dx \Rightarrow \int \frac{df}{f} = \int \lambda dx + c \quad \text{(any constant)}$$

$$\Rightarrow \log f = \lambda x + c \Rightarrow f = e^{\lambda x} e^{c} \quad \text{(call it a)}$$

$$\Rightarrow \underline{f = a e^{\lambda x}}$$

It shows that for every $\lambda \in \mathbb{R}$, we have functions which are multiples of $e^{\lambda x}$ such that $\underline{\lambda \text{ is eigenvalue}}$ & $\underline{e^{\lambda x} \text{ is eigenvector}}$.

Lookup <u>modes of vibration</u> of musical instruments (membranes). All of them correspond to eigenvectors of a <u>suitable</u> differential operators!
|||ly modes of vibration of civil engg. structures etc.

eg 3 The famous PCA (Principal component analysis) technique used for efficiently representing data also solves a suitable e.v.p.! In fact, the state-of-the-art face recognition algorithms employ this technique to figure out the "best" representation of images containing faces.

Lookup and read articles on eigenfaces.

In the above, we just tried to list down some various appl. in which evp appear & note the diversity in the appl. fields! Hence it is important to study eigenvalue theory in detail.

Now a solution exists for evp iff $\exists\, x, \lambda \ni$

$$Mx = \lambda x \ , \ x \neq 0$$

$$\iff \lambda x - Mx = 0, \ x \neq 0$$

In class, we said $Mx - \lambda x = 0,\ x \neq 0$. Both these are the same

identity matrix of suitable size

$$\iff (\lambda I - M)x = 0, \ x \neq 0$$

$$\iff \underbrace{\dim \overbrace{N(\lambda I - M)}}_{} \geq 1 \text{ for some } \lambda$$

known as eigenspace of the eigenvalue λ.

null space

this is known as the geometric multiplicity of the eigenvalue λ

$$\iff \dim C(\lambda I - M) < n$$

In other words eigenvectors corresponding to λ are vectors in eigen space

$$\iff (\lambda I - M) \text{ is not an invertible matrix}$$

Recall that $M^{-1} = \frac{1}{\det M} C^T$ (transpose of matrix of cofac.)

i.e. M^{-1} exists $\iff \det M \neq 0$

$$\iff \det(\lambda I - M) = 0 \quad \text{for some } \lambda \quad \text{(I)}$$

Now it is easy to see that if we expand $\det(\lambda I - M)$ we get a polynomial in λ! (of degree n) why?

eg let $M = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$ $\quad \det(\lambda I - M) = \det\left(\begin{bmatrix} \lambda-2 & 0 \\ 0 & \lambda-3 \end{bmatrix}\right) = \underbrace{\lambda^2 - 5\lambda + 6}_{\text{quadratic in } \lambda}$

$|||$ly for $M_{n\times n}$ we get a polynomial in λ $\longrightarrow$ leading coeff. of λ^n will be 1.

~~Eigen vo~~ As per I, the eigen values are nothing but the roots of this polynomial! This polynomial is known as the characteristic polynomial of the matrix M. In other words, eigen values are roots of the characteristic polynomial of M.

Note that, characteristic polynomial & eigenvalues are hence properties of the matrix M i.e. given M, these get fixed. However the eigenvector can be ~~des~~ described given both M & λ! Hence eigen vector is a property of ~~both~~ the Matrix, eigenvalue pair.

In the above eg. the roots of the ch. poly. are 2 and 3 respectively & hence the eigen values are 2 and 3.

Lets known give a convincing answer to the question ~~of~~ about existence of eigen values:

~~Ea~~ e.v. are roots of ch. poly $\rightarrow$ which is nothing but polynomial of degree n

$\therefore$ By fundamental theorem of algebra, we have, that there will be exactly 'n' roots (if all complex & repeated roots are counted).

Moreover, since the coefficients of this polynomial are real, the ~~roots~~ complex roots can only occur in pairs (of conjugates!)

$\Rightarrow$ If dim. of M i.e. n is odd, then for sure, at least one eigenvalue is real (for us e.v. real means e.v. exists) since we are concerned with real vec. spa. only

However for a matrix of even dimension, it may happen that all eigenvalues are complex (i.e. there are no eigenvalues)

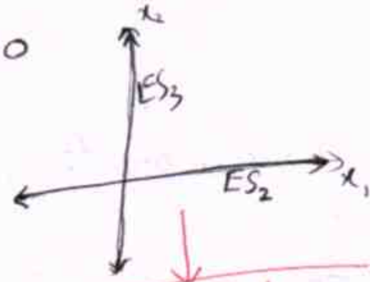
Now lets look at some more eg. which will give us some more insight.
Lets calc. eign. values & corresponding eigenspaces for various 2x2 matrices:

eg a for the matrix $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$ we already know that eig. values of 2 & 3.

Now eigenspace of 2 is null space of $2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$

$x \in ES_2 \iff \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \iff -x_2 = 0 \iff x_2 = 0$

i.e. ES_2 is nothing but x_1-axis!



only orientations invariant to M are x_1-axis & x_2-axis

$x \in ES_3 \iff \left(3\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\iff x_1 = 0$ i.e. the x_2-axis!

Note that ES_2 & ES_3 together span the entire $\mathbb{R}^2$!

eg b $M = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ $\quad \det(\lambda I - M) = \det\left(\lambda\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right) = \det\left(\begin{bmatrix} \lambda-2 & 0 \\ 0 & \lambda-2 \end{bmatrix}\right)$
$= \underbrace{(\lambda-2)(\lambda-2)}_{\text{ch. polynomial of } M}$.

roots of ch. poly. are $\underbrace{2 \;\&\; 2}$ $\rightarrow$ repeated eigenvalue!

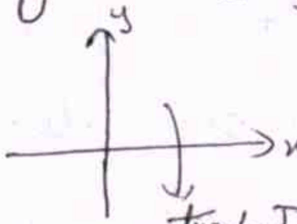
No. of times an eigenvalue is root of the ch. poly is known as the arithmatic multiplicity of the eigen. value!

Now, eigenspace of the eig. val. 2 is given by null space of $2\begin{bmatrix}1&0\\0&1\end{bmatrix}-\begin{bmatrix}2&0\\0&2\end{bmatrix}=\begin{bmatrix}0&0\\0&0\end{bmatrix}$

— which is entire $\mathbb{R}^2$!

$\therefore$ Take any ~~direction~~ orientation in $\mathbb{R}^2$ it is ~~form~~ invariant w.r.t. the transformation $\begin{bmatrix}2&0\\0&2\end{bmatrix}$

(realize that geometrically also this is true)



Two questions arise from these examples:

any two $\perp$ vec here form a basis for eigenspace of 2!

(i) it looks like eig. val. of a diagonal matrix are the diagonal ~~matrix~~ entries itself. Is it true? <u>Yes</u> (why?)

(ii) it looks like g.m. = a.m. for an eigenvalue. Is it true. **NO**

g.m.: geometric multiplicity; a.m.: arithmetic multiplicity

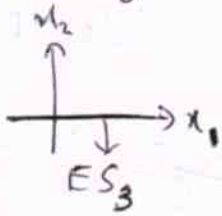
here is counter example:

eg^c $\quad M = \begin{bmatrix}3 & a\\ 0 & 3\end{bmatrix}$. (some $a \in \mathbb{R}, a \neq 0$) ch. poly. is $(\lambda-3)^2=0$ $\therefore$ eig values are 3, 3

a.m. = 2 again

However eigenspace of eig. val. 3 is null space of $\begin{bmatrix}0 & -a\\ 0 & 0\end{bmatrix}$

which is the x_1-axis



$\Rightarrow$ g.m. = 1 $\neq$ a.m. = 2

Therefore g.m. may not be equal to a.m. However one can easily see that g.m. $\leq$ a.m. (why?)

Proof by contradiction:

egc also raises the question:

→ is it true that the eig. val. of an upper triangular matrix, are again the entries on the diagonal? Yes! why?

→ ||ly for a lower triangular matrix.

For diagonal, upper Δ, lower Δ matrices, the eigenvalues are the entries on the principal diagonal of the matrix.

egc also showed us that in cases g.m. < a.m., then the eigen spaces will not span the entire $\mathbb{R}^n$.

ega & b showed that g.m = a.m. then eigenspaces together span $\mathbb{R}^n$

So are these statements true ?? Yes (here is the theorem & proof)

Theorem: Let $\lambda_1, \lambda_2, \ldots, \lambda_m$ be distinct eigenvalues of a $M_{n\times n}$ matrix. Choose any eigenvector v_1 from eigenspace of λ_1

" v_2 " " of λ_2

⋮

" v_m " " of λ_m.

Then $\{v_1, v_2, \ldots, v_m\}$ are linearly independent!

Proof (by contradiction)

Suppose $v_1, v_2, \ldots, v_m$ are linearly dependent. Then I can express at least one vector say v_j as linear combination of others.

(since all are non-zero vectors)

i.e. $v_j \in \text{span}(\{v_1, v_2, \ldots, v_{j-1}, v_{j+1}, \ldots, v_m\})$

~~Now if we keep throwing away vectors for~~
~~we can throw away~~

Now if this set of vectors is LI then fine, otherwise we can throw away one vector (which can be expressed as lineal comb of others) and the span of that reduced set will be the same. Illly I can keep throwing vectors till we have a LI set having same span as

Without loss of generality, let us assume $v_1, v_2, \ldots, v_n$ are LI & span

i.e. $v_j = \alpha_1 v_1 + \cdots + \alpha_n v_n$ where $v_1, v_2, \ldots v_n$ are LI

(Of course all $v_j, v_1, \ldots v_n$ are eigenvectors & hence non-zero)

Now lets multiply both sides of above eqn. by M:

$$M v_j = \alpha_1 M v_1 + \cdots + \alpha_n M v_n$$

$$\Rightarrow \lambda_j v_j = \alpha_1 \lambda_1 v_1 + \cdots + \alpha_n \lambda_n v_n$$

$$\Rightarrow \lambda_j(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 \lambda_1 v_1 + \cdots + \alpha_n \lambda_n v_n$$

$$\Rightarrow \alpha_1(\lambda_1 - \lambda_j)v_1 + \alpha_2(\lambda_2 - \lambda_j)v_2 + \cdots + \alpha_n(\lambda_n - \lambda_j)v_n = 0$$

$$\Rightarrow \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0 \quad \left(\text{since } v_1, v_2, \ldots, v_n \text{ are LI } \& \ \lambda_j \neq \lambda_i,\ j \neq i\right)$$

$\Rightarrow v_j = 0$ which is not possible as v_j is an eigen vector!

Hence Proved.

The above theorem answers our question.

Suppose there are 'n' distinct eigenvalues for M. then by theorem we can pick 'n' eigenvectors which span entire $\mathbb{R}^n$!

Or

In general, if there are 'n' distinct eigenvalues for M, and few of them are repeated eig.val., then as long as a.m. = g.m. for each of those eigenvalues, we will be able to span entire $\mathbb{R}^n$! (why?)

We will take this setting as our "role model" and make further arguments. (we can repeat with this setting also)

(it is easy to see that in this case a.m. = g.m. = 1 for all the eigenvalues!)

Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be distinct eigenvalues of a matrix $M_{n\times n}$ and
$v_1, v_2, \ldots, v_n$ be some set of eigenvectors (one per eigenvalue)

Now one can represent the equations
$$\left.\begin{aligned} Mv_1 &= \lambda_1 v_1 \\ &\vdots \\ Mv_n &= \lambda_n v_n \end{aligned}\right\} \text{set of n vector eqns. each representing n scalar equations}$$
in a compact form (through a single matrix equation)

$$\boxed{MX = X\Lambda}$$
(why?)

represents n^2 scalar equations or equivalently represents n vector equations

Here X is the matrix with i^{th} column as v_i
Λ is the diagonal matrix with entries as $\lambda_1, \lambda_2, \ldots, \lambda_n$.

Now X is a matrix where all columns are LI (by the theorem).

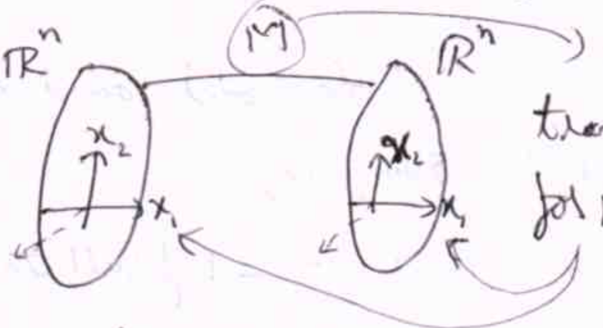
Hence it is invertible!

<u>i.e.</u> $MX = X\Lambda \iff \boxed{M = X\Lambda X^{-1}}$

this particular decomposition of the matrix is called the eigen value decomposition (EVD) of the matrix.

This EVD is what ~~[illegible]~~ helps us diagonalize the matrix M! because:

~~Suppose~~



we know M represents the transformation matrix when the std. basis for $\mathbb{R}^n$ is chosen

Now suppose instead of choosing the std. basis of M, we chose the following basis $\{v_1, v_2, \dots, v_n\}$ → (in the case with distinct eigvec. they are LI)

It is easy to the transformation that links the original basis & the new basis is:

$$\underbrace{\{x_1, x_2, \dots, x_n\}}_{\text{std. basis}} \underset{X}{\overset{X^{-1}}{\rightleftarrows}} \underbrace{\{v_1, v_2, \dots, v_n\}}_{\text{basis of eigen vectors}}$$

<u>why?</u> → here's why ~~[illegible]~~

lets show 1-1 correspondance between axis (basis vectors) here and here through the matrix X:

What are the coordinates of v_1 vector under the basis $\{v_1, v_2, \ldots, v_n\}$?

~~Ho~~ it is ~~on~~ nothing but $\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

But ~~(H)~~ the coordinates of v_1 vector ~~is~~ under the std. basis is given by v_1 itself.

of course $\quad v_1 = \underbrace{X \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\text{only first column in } X \text{ which is } v_1 \text{ is picked}} \quad$ also $\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = X^{-1} v_1$

|||ly for any eigenvector v_i we have $\quad v_i = X \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \rightarrow i^{th}$ position

$\&\ \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = X^{-1} v_i$

Now suppose the 1-1 linear transformation X^{-1} is applied & the basis is changed to $\{v_1, v_2, \ldots, v_n\}$. Then let us see how the transformation matrix M would look like:

*[Recall ~~to~~ how we used to write down transformation matrix of a given linear transformation]

Under the new basis, ~~M the~~ the new tra. matrix would be nothing but:

$$\begin{array}{c} \begin{array}{ccc} M\cancel{v_1} & Mv_2 & \cdots\ Mv_n \\ \downarrow & \downarrow & \cdots\ \downarrow \end{array} \\ \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & \vdots \\ \vdots & 0 & \cdots & \vdots \\ \vdots & \vdots & & 0 \\ 0 & 0 & & \lambda_n \end{bmatrix} \end{array} \longrightarrow \text{because } Mv_i = \lambda_i v_i$$

Hence we have diagonalized the matrix!

Here's another way to see what is happening:

Suppose in the original basis a vector v is taken to vector w by M

i.e. say $Mv = w$.

Now we got that $M = X\Lambda X^{-1}$ is the evd.

i.e. $X\Lambda X^{-1} v = w \iff \Lambda (X^{-1}v) = (X^{-1}w)$

$$\therefore \boxed{Mv = w} \iff \boxed{\Lambda v' = w'}$$

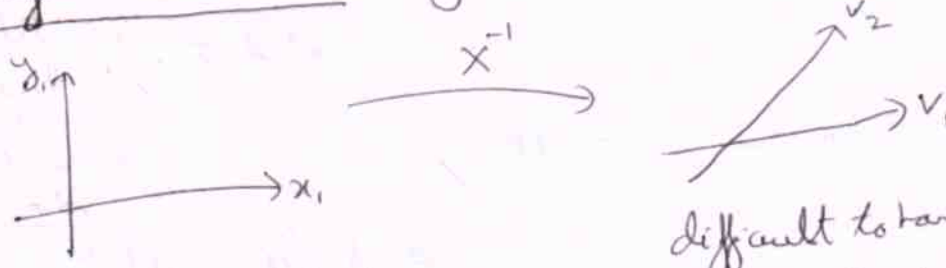
where $v' = X^{-1}v$ & $v = Xv'$ } i.e. if the basis is chosen to be $\{v_1, v_2, \ldots, v_n\}$ rather than the std. basis!

Now if any matrix is decomposable as $M = X\Lambda X^{-1}$, ~~then~~ (irrespective of whether eigenvalues of M are distinct or not) then also all the above arguments will hold and the matrix will be diagonalizable. ~~Also~~ Also the values in the diagonal matrix Λ will be the eigenvalues & the column vectors of X will be eigen vectors corresponding to them.

* [of course this decomposition is not unique. i.e. X is not unique. We are free to choose and LI eigenvectors]

All said and done though we were able to diagonalize the matrix M still we would have done it only when the original ~~&~~ std. basis (which is nice because it is an orthonormal basis of $\mathbb{R}^n$!) is replaced with new basis $\{v_1, v_2, \ldots, v_n\}$

Of course this new basis vectors $\{v_1, \dots, v_n\}$ i.e. eigenvectors are LI but may not be orthogonal!



nice orthonormal basis

difficult to handle generic basis of $\mathbb{R}^n$!

So this raises the important question: Can a matrix be diagonalized such that is again orthonormal basis? (If so it would be very convinient). The answer to this question is given by Spectral Theorem. which says in case of symmetric matrices this always can be done & moreover symmetric matrices are the only matrices where this can be done! Before going into the statement & proof of Spectral theorem, let us just explore some prop. of eig. val. & uses of diagonalizing a matrix:

Firstly $\det(A) = \det(A^T)$
$\therefore \det(\lambda I - M) = \det((\lambda I - M)^T)$
$= \det(\lambda I - M^T)$
Hence ch. poly. of M & M^T are the same!

We know

$$\underbrace{\det(\lambda I - M)}_{\text{ch. poly.}} = \underbrace{(\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)}_{\text{factorizes like this if } \lambda_1, \lambda_2, \dots, \lambda_n \text{ are the eigenvalues.}}$$

$\Rightarrow$ eig. values of M & M^T are same.
However eig. vec. will be different in general.

Now put $\lambda = 0$ in the above equation, we get $\det(-M) = (-1)^n \det(M)$
$= (-1)^n \lambda_1 \lambda_2 \cdots \lambda_n$

In other words,

$$\boxed{\det M = \text{prod. of its eigen values}}$$

this says $\det(M) \neq 0 \Leftrightarrow$ none of eig values of M are zero. Hence this can be used as a way to characterizing an invertible matrix

(note that this is valid even if eig. val. are complex or repeated etc.)

Matrix is invertible $\Leftrightarrow$ none of eig. val. are zero

IIIly we know from algebra that if

$\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0 = 0$ is a polynomial then, the sum of roots of the poly. is $-a_{n-1}$ i.e. $-$coeff. of λ^{n-1}

Hence the sum of eigenvalues will be the $-$coeff. of λ^{n-1} in the characteristic polynomial!

Now a little thought must convince that the only term in the expansion of the determinant

$$\det(\lambda I - M) = \det\left(\begin{bmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & & & \\ -a_{n1} & \cdots & \cdots & \lambda - a_{nn} \end{bmatrix}\right)$$

which can give a non-zero coeff. for λ^{n-1} is the following term:

$$(\lambda - a_{11})(\lambda - a_{22}) \cdots (\lambda - a_{nn}) \qquad (\text{why?})$$

The coeff of λ^{n-1} in ? is $-(a_{11} + a_{22} + \cdots + a_{nn}) = -\text{trace}(M)$

Hence $\boxed{\text{sum of eigen values} = \text{trace}(M)}$

Now, knowledge of prop. like these will enable to us write down eig.val. of a matrix without actually having to determine roots of the char. poly.! Lets do some examples:

eg $M = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Lets calc. eig. val. without looking at the ch. poly.

We know $\det(M) = 0$ & $\text{trace}(M) = 2$

$\Rightarrow$ atleast one of eig. values must be zero; sum of the eig. values is 2

$\therefore$ the eigenvalues are $\underline{0, 2}$

eg $M = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$. Here also $\det(M) = 0$ & $\text{trace}(M) = 3$

so atleast one eigen value is zero; sum of eig. val. is 3.

So the eig. values must be of the form $0, e, 3-e$. we need to still determine e. We can do that by looking at g.m. of the eigenvalue 0!

$$\text{g.m. of } 0 = \dim N\Big(\overbrace{\begin{bmatrix} -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{bmatrix}}^{\text{since rank}=1}\Big) = 2 \leq \text{a.m. of } 0$$

but max. possible value of a.m. of 0 is 2

$\therefore$ g.m. of 0 = a.m. of 0 $= 2$ in this case.

Here the eigen values are $\underline{0, 0, 2}$

IIIly if $M = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$ i.e. $n \times n$ matrix of all ones!

it is easy to see that eigen values will be $\underbrace{0, 0, \ldots, 0}_{n-1 \text{ times}}, n$

*[Imagine doing this by determining roots of ch. poly. of M!]

Hence there are powerfull properties & worth remembering!

Let us now look at some adv. of diagonlzng of a matrix.

i.e. $\underline{M = X \Lambda X^{-1}}$

First of all computing powers of M is easy: (recall that M^2 represents composition of transformation)

$$M^2 = MM = (X\Lambda \underbrace{X^{-1})(X}_{I}\Lambda X^{-1}) = X\Lambda^2 X^{-1}$$

|||ly $\underline{M^p = X\Lambda^p X^{-1}}$ for any $p \in \mathbb{N}$

Computing this is computationally intensive

computing Λ^p is easy! just need to raise each diagonal entry to power p.

Imagine an application where we need to compute various powers of M clearly using $X\Lambda^p X^{-1}$ is more viable (even if we include computations reqd. for performing the evd in the first place!)

It is not hard to think abt such appl. In fact we have studied one such scenario in this course!

Take eg. of a finite state Markov chain then

$$\boxed{\pi_n^T = \pi_0^T P^n}$$

prob. dist. at n^{th} stage

initial prob. dist.

trans. prob. matrix

in this case we would like to compute various powers of P and doing it via the evd of P (as indicated above) is more viable option!

~~In fact ...~~

~~Also~~ Now suppose M is invertible i.e. none of eig. val. are zero then, we claim $M^{-1} = X\Lambda^{-1}X^{-1}$

<u>why?</u> because $M^{-1}M = X\Lambda^{-1}X^{-1}X\Lambda X^{-1}$
$= X\Lambda^{-1}\Lambda X^{-1}$
$= XX^{-1} = I$

$MM^{-1} = X\Lambda X^{-1}X\Lambda^{-1}X^{-1}$
$= X\Lambda\Lambda^{-1}X^{-1}$
$= XX^{-1} = I$

$\therefore \boxed{M^{-1} = X\Lambda^{-1}X^{-1}}$ ||ly $\boxed{M^p = X\Lambda^p X^{-1}}$ where $p \in \mathbb{Z}$

computing is difficult — computing is very easy

Hence raising a matrix to any integral power is very easy.
In fact writing down a polynomial in M is itself very easy:

$$\underbrace{a_0 I + a_1 M + \dots + a_m M^m}_{\text{polynomial in } M} = a_0 \overbrace{XX^{-1}}^{I} + a_1 X\Lambda X^{-1} + \dots + a_m X\Lambda^m X^{-1}$$
$$= X\left(a_0 I + a_1\Lambda + \dots + a_m\Lambda^m\right)X^{-1}$$

difficult to compute

same polynomial in Λ

easy to compute

$$a_0 I + a_1\Lambda + \dots + a_m\Lambda^m = \begin{bmatrix} a_0 + a_1\lambda_1 + \dots + a_m\lambda_1^m & & 0 \\ & \ddots & \\ 0 & & a_0 + a_1\lambda_n + \dots + a_m\lambda_n^m \end{bmatrix}$$

apply same polynomial to each eigen value!

$\therefore$ A polynomial of a matrix (which is itself a matrix) has eigen vectors same as those of M. Moreover the eigen values can be obtained using eigen values of M easily!

$$\therefore \quad \boxed{p(M) = X\,p(\Lambda)\,X^{-1}}$$

$\underbrace{p(M)}_{\text{polynomial in } M}$ $\quad$ $\underbrace{p(\Lambda)}_{\text{same poly. in } \Lambda}$

Now we know analytic functions are "natural" extensions of polynomial (to infinite degree). We can show (not in this course) that

$$\boxed{f(M) = X\,f(\Lambda)\,X^{-1}} \quad \text{I}$$

$\underbrace{f(M)}_{\text{analytic function of } M}$ $\quad$ $\underbrace{f(\Lambda)}_{\text{same analytic fn. of } \Lambda}$

again adv. is this is easy to compute.

eg Suppose we want to compute the matrix e^M.
(recall that e^x is an analytic function!)

the default way is to compute:

$$e^M = I + M + \frac{M^2}{2!} + \cdots + \frac{M^n}{n!} + \cdots$$

which is very difficult to do.

but using I $\quad e^M = X\,\underbrace{e^\Lambda}\,X^{-1}$ nothing but → (why?) hence easy to compute!

$$\begin{bmatrix} e^{\lambda_1} & & & \\ & e^{\lambda_2} & & 0 \\ & & \ddots & \\ 0 & & & e^{\lambda_n} \end{bmatrix}$$

I ~~represents~~ says that some particular transformations of M (i.e. analytic func. of M) do not effect eigen vectors but eigenvalues are affected (of course the corresp. eig val. can be easily computed using the original evals of M)

One can talk about transformations of M where eigen values are preserved but eigen vectors are not!

Such transformations are known as similarity transformations:

given an invertible matrix S look at the following transformation of matrix M: SMS^{-1}. It is easy to show that SMS^{-1} has eig val same as that of M!

$$SMS^{-1} = S\underbrace{X\Lambda X^{-1}}_{\text{evd of } M}S^{-1} = Y\Lambda Y^{-1} \quad \text{II} \longrightarrow Y = SX\text{, say}$$

This says eigenvalues of SMS^{-1} are also given by Λ. ~~So~~ However eigenvectors of SMS^{-1} are not the same as those of M!.

Actually ideas presented in I, II are used in some of the algorithms to compute evd of a given matrix.

Unfortunately we do not have time to discuss algorithms useful for performing evd, however we list them below:

→ Spectral Theorem

Theorem M is symmetric $\iff$ $M = X\Lambda X^T$ where $X_{n\times n}$ is orthonormal & $\Lambda_{n\times n}$ is diagonal

Before proving this, lets understand the statement: X is orthonormal $\Rightarrow X^T = X^{-1}$. Hence $M = X\Lambda X^T$ is nothing but the evd of M! Here entries in Λ are eigenvalues & columns of X are the corresponding eigen vectors. X is orthonormal $\Rightarrow$ eigenvectors are orthogonal!
i.e. eigen spaces are orthogonal!

Further by our previous analysis, the matrix will be diagonal if the basis is taken as $\{v_1, v_2, \ldots, v_n\}$ but now these form an orthonormal basis!

Hence the statement of Spectral thm can also be re-written as:

Theorem: Symm. matrices are the only matrices which can be diagonalized using an orthonormal transformation of original vectors & for all symm matrices ~~the~~ such a diagonalization is possible.

Proof $\quad$ TST $\quad M$ is symm $\Longleftrightarrow M = X \Lambda X^T$

orthonormal

diagonal.

Proof of $\Leftarrow$ is easy: $M^T = (X\Lambda X^T)^T = X\Lambda X^T = M$. Hence Proved

Proof of $\Rightarrow$: Lets do the proof in steps:

(i) Lets prove all eigenvalues of symm. matrix are real. (this is necessary)

(ii) the eigenspaces corresponding to distinct eigenvalues are orthogonal

(iii) Always can choose ~~all~~ 'n' orthogonal eigenvectors of M.

(Actually we can skip (ii) if (iii) is proved, but gives more insight so lets do it)

Convince yourself that if (i), (iii) are proved then $M = X\Lambda X^T$.

Proof of (i): We will prove a stronger statement:

If M is Hermetian, then all eigenvalues ~~of~~ are real.

~~&~~ M is Hermetian iff $M = \bar{M}^T$ ($\bar{M}$ is conjugate of M i.e. matrix got by taking conjugate of each element of M)

$\bar{M}^T$ is denoted by M^*

eg. Let $M = \begin{bmatrix} a_{11}+ib_{11} & a_{12}+ib_{12} \\ a_{21}+ib_{21} & a_{22}+ib_{22} \end{bmatrix} \rightarrow$ generic 2x2 complex matrix

$$\overline{M} = \begin{bmatrix} a_{11}-ib_{11} & a_{12}-ib_{12} \\ a_{21}-ib_{21} & a_{22}-ib_{22} \end{bmatrix} ; \quad \overline{M}^T = \begin{bmatrix} a_{11}-ib_{11} & a_{21}-ib_{21} \\ a_{12}-ib_{12} & a_{22}-ib_{22} \end{bmatrix}$$

Now $M = \overline{M}^T$ iff $a_{11}+ib_{11} = a_{11}-ib_{11} \Leftrightarrow b_{11}=0$ i.e. entry is real

$a_{22}+ib_{22} = a_{22}-ib_{22} \Leftrightarrow b_{22}=0$ i.e. entry is real

$\left.\begin{array}{l} a_{12}+ib_{12} = a_{21}-ib_{21} \\ a_{21}+ib_{21} = a_{12}-ib_{12} \end{array}\right\}$ off diagonals are conjugates of each other!

Hence Hermetian matrix is a matrix (with possibly complex entries) where all diagonal entries are reals and off diagonals entry ij is conjugate of ji^{th} entry.

Now it is easy to see that if M is Hermetian & M is real, then M is symmetric.

$\therefore$ All symmetric matrices are Hermetian!!

Hence proving all eigen values of Hermetian matrix are real will indeed prove (i)!

*[Recall that complex matrices represent lin tra. in complex vector spaces. Also we noted earlier, complex vector spaces are easier to handle than real vec. spa.]

Now M is a complex matrix it will have n complex eigen values (some may be repeated). Take any one eigen value λ. We have:

$$Mx = \lambda x, \quad x \neq 0 \quad (a)$$ (if λ is eig. val., then $\exists\, x$ satisfying the evp)

Now take * on both sides,

$$(Mx)^* = (\lambda x)^*, \quad x \neq 0$$ (* is analogous to transpose operation on real matrices & vectors)

$$\Rightarrow x^* M^* = \bar{\lambda} x^*, \quad x \neq 0 \quad (b)$$
$$\Rightarrow x^* M = \bar{\lambda} x^* \longrightarrow$$ because $M = M^*$ (M is Hermetian)

Lets ~~~~ pre-multiply (a) by x^* & post multiply (b) by x

we get: $$x^* M x = \lambda x^* x$$
$$x^* M x = \bar{\lambda} x^* x$$
$$\Rightarrow (\lambda - \bar{\lambda}) x^* x = 0$$

$\Rightarrow$ either $\lambda = \bar{\lambda}$ or $x^* x = 0$

Now let $x = \begin{bmatrix} a_1 + ib_1 \\ \vdots \\ a_n + ib_n \end{bmatrix} \rightarrow$ generic n dim complex vector.

$$x^* x = \begin{bmatrix} a_1 - ib_1 & \cdots & a_n - ib_n \end{bmatrix} \begin{bmatrix} a_1 + ib_1 \\ \vdots \\ a_n + ib_n \end{bmatrix} = (a_1^2 + b_1^2) + \cdots + (a_n^2 + b_n^2)$$

This says $x^* x \geq 0 \;\forall x$ & $x^* x = 0 \Rightarrow x = 0$

But x is eigenvector hence $x \neq 0$ $\therefore$ $\lambda = \bar{\lambda}$ i.e. λ is real.

Of course this happens with every eigenvalue of M. Hence Proved.

*[Note that we can't repeat same proof explicitly taking M is symm. as then we would not gauranteed to have atleast one real eig. val. — with this proof we have cleverly avoided this complication!]

Proof of (ii): Again we will prove eigenspaces of Hermetian matrix are orthogonal!

Suppose $\lambda_1 \neq \lambda_2$ are two distinct eigenvalues of M, then:

$$M x_1 = \lambda_1 x_1 \quad , x_1 \neq 0$$
$$M x_2 = \lambda_2 x_2 \quad , x_2 \neq 0.$$

(x_i is eigenvector corresponding to λ_i)

Now take $*$ on both sides of first eqn. & post multiply by x_2 we get

$$x_1^* M x_2 = \lambda_1 x_1^* x_2$$

Premultiply second eqn. by x_1^*, we get:

$$x_1^* M x_2 = \lambda_2 x_1^* x_2$$

$$\Rightarrow (\lambda_1 - \lambda_2) x_1^* x_2 = 0$$
$$\Rightarrow x_1^* x_2 = 0$$

Now in complex vector space $x_1^* x_2$ represents inner product! It is easy to verify all three non-negativity (already we showed above), symmetry & linearity axioms hold for $*$

Hence $x_1^* x_2 = 0 \Rightarrow x_1$ & x_2 are orthogonal! Hence Proved!

Proof of (iii) This is by induction, here we will give the sketch of proof:

Pick any eigenvector of M say $\lambda_1 \neq 0$ & corresponding eigenvector v_1

Now let S be the orthogonal complement of subspace spanned by v_1

$$\underset{\dim = n-1}{S} \qquad \underset{\dim = 1}{v_1}$$

i.e. $v \in S \Leftrightarrow v \perp v_1$

Now it is interesting to observe that $Mv \perp v_1$!

$$Mv \perp v_1 \Leftrightarrow v_1^T M v = 0$$
$$\Leftrightarrow v^T M v_1 = 0 \quad (\because \underline{M = M^T})$$

$\Leftrightarrow v^T \lambda_1 v_1 = 0$

$\Leftrightarrow v^T v_1 = 0 \qquad (\because \lambda_1 \neq 0)$

* [Note that we can always pick $\lambda_1 \neq 0$ otherwise if all the eigen values of a symmetric matrix are zero then the matrix itself is ~~a~~ the zero matrix] *

So vectors in orthogonal complement of M are mapped to itself by M! Now we can view $M: S \to S$ (instead of $M: \mathbb{R}^n \to \mathbb{R}^n$) again M is symm. but $\dim S = n-1$.

Now we are back at step 1! we can repeat this business until there are no more non-zero eigen values! Hence we will get orthogonal vectors

$v_1, v_2, \ldots, v_r$ where r is rank of matrix M.

(why?) because zero eigenvalues are nothing but vectors in Null space of M itself!

Now we can take any orthogonal basis of Null space of M and complete $\{\underbrace{v_1, v_2, \ldots, v_r}, \underbrace{v_{r+1}, v_{r+2}, \ldots v_n}\}$

orthogonal basis of column space or row space got by above method

also are the eigen vectors corresponding to non zero eigen values

any basis for null space of M is OK

eigen vectors corresponding to the zero eigen value.

Hence (iii) is Proved!

Once spectral theorem is in place it is easy to show that:

$$M \text{ is pnd} \iff M = \underbrace{X}_{\text{orthonormal}} \overbrace{\Lambda}^{\text{diagonal with entries} \geq 0} X^T$$

See practice problems

$$\text{IIIly } M \text{ is pd} \iff M = \underbrace{X}_{\text{orthonormal}} \overbrace{\Lambda}^{\text{diagonal with entries} > 0} X^T$$

Once these are proved we have a new way of ~~classifying whether a matrix~~ categorizing matrices depending on their eigenvalues:

if all eigvals ~~of~~ are real then M is symmetric
if " " are ≥ 0 then M is pnd
if " " " > 0 then M is pd.

eg for any rectangular $M_{m\times n}$ we know that both M^TM and MM^T are pnd. $\therefore$ eigenvalues of MM^T and M^TM are all ≥ 0

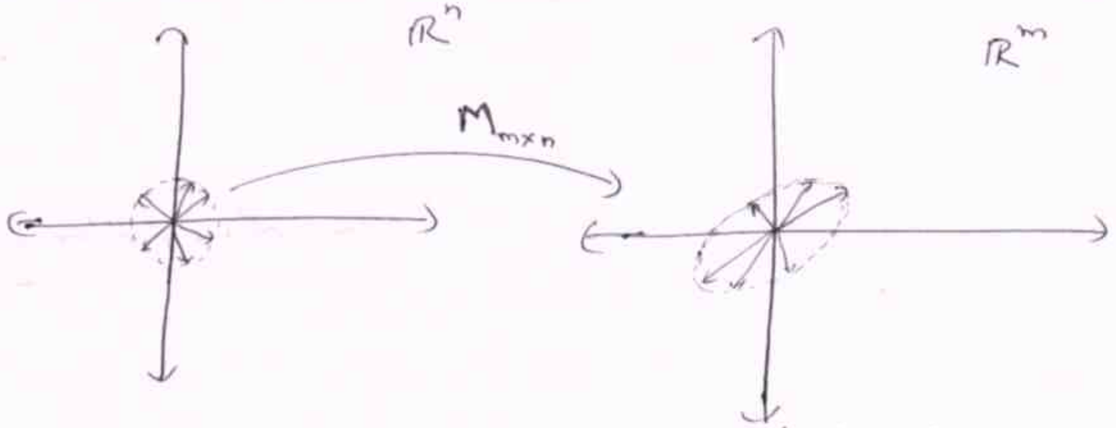
Also if M has full column rank then we know M^TM is invertible
$\therefore M^TM$ is pd $\Rightarrow$ all e.val. are > 0

if M has full row rank, then MM^T is invertible
$\Rightarrow MM^T$ is pd $\Rightarrow$ all e.val. are > 0.

$\rightarrow$ In the following we will give a new interpretation to the notion of eigen~~va~~vector which we have not imagined so far.

Consider the set of all vectors in $\mathbb{R}^n$ with unit length. (locus of all these pts. is the std. unit circle at origin). In other words we are considering all directions in $\mathbb{R}^n$. ~~we will also~~ Suppose now we apply trans. rep. by M

Here is picture in 2-d. (projection is 2-d)



We will show that the directions which gets maximally scaled are nothing but the eigen vectors corresponding to the ~~...~~ eig. value of M^TM with maximum magnitude!!

In other words we will show

$$\left.\begin{array}{ll}\max\limits_{x\in\mathbb{R}^n} & \|Mx\| \\ \text{s.t.} & \|x\|=1\end{array}\right\} = \sqrt{\lambda_{max}} \longrightarrow \text{eigenvalue of } M^TM \text{ with maximum } \cancel{\text{absolute}} \text{ value}$$

$$\left.\begin{array}{ll}\operatorname{argmax}\limits_{x\in\mathbb{R}^n} & \|Mx\| \\ \text{s.t.} & \|x\|=1\end{array}\right\} = \cancel{\text{eigenvector}} \text{ unit vector in eigenspace corresponding to}$$

$\boxed{x\in\mathbb{R}^n \ \& \ Mx\in\mathbb{R}^m}$

Now the objective $\cdot$ $\|Mx\| = \sqrt{x^TM^TMx}$

$= \sqrt{x^TX\Lambda X^Tx}$ where $X\Lambda X^T$ is evd of M^TM (we know ≥ 0)

$= \sqrt{y^T\Lambda y} \longrightarrow$ let $X^Tx = y$

$= \sqrt{\lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2} \longrightarrow$ let $\Lambda = \begin{bmatrix}\lambda_1 & & & 0\\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n\end{bmatrix}$

let $y = \begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{bmatrix}$

Also $\|x\|=1 \iff \underbrace{\sqrt{x^Tx}=1}_{x=Xy} \iff \sqrt{y^TX^TXy}=1 \iff \sqrt{y^Ty}=1 \iff \|y\|=1$

$\iff \|y\|^2=1$

Hence using the 1-1 orthonormal transformation $X^T x = y$, we can rewrite the original optimization problem in terms of y:

$$\max_{y \in \mathbb{R}^n} \sqrt{\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n}$$

$$\text{s.t.} \quad \sqrt{y_1^2 + y_2^2 + \cdots + y_n^2} = 1$$

This is very easy to solve compared to original problem (~~esp~~) (especially for CS students!) (why?) : (explanation was given in lecture)

The optimal is achieved when $y = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ → entry corresponding to λ with maximum value.

Suppose we order the eigen values in decreasing order of magnitude : $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

then $y = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ will be one choice of y where the objective is maximized!

& so the ~~ob~~ maximized objective $= \sqrt{\lambda_1}$

$\because x = Xy$

Now the x corresponding to $y = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ is nothing but v_1, which is nothing but ~~an~~ an eigen vector in eigenspace of eigen value with maximum magnitude. Hence Proved.

In the 2-d fig. above, the dir. of major axis of ellipse will be that of first eigen vector of $M^T M$.

On a similar note, the "second eigenvector" will be direction corresponding to the second largest scaling & so on.. In fact that corresponding to the "least eigenvector" will be minimally scaled.

Now a similar question can be asked about the matrix $M^T_{n\times m}$

$$\max_{x\in\mathbb{R}^m} \|M^Tx\| \quad \text{s.t.} \quad \|x\|=1$$

$\rightarrow$ obviously the optimum will be $\sqrt{\bar{\lambda}_1}$ where $\bar{\lambda}_1$ is the maximum eig. val. of MM^T.

In fact we can show $\sqrt{\bar{\lambda}_1} = \sqrt{\lambda_1}$! i.e. eigen values of M^TM and MM^T are exactly the same & this common value $\sqrt{\lambda_1}$ is called the first singular value of the rectangular matrix M!

In fact there is a neat relation between

$$\operatorname*{argmax}_{x\in\mathbb{R}^n} \|Mx\| \quad \underline{\underline{\text{s.t.}}} \quad \|x\|=1 \qquad \text{and} \qquad \operatorname*{argmax}_{x\in\mathbb{R}^m} \|M^Tx\| \quad \underline{\underline{\text{s.t.}}} \quad \|x\|=1$$

and leads to the concept of singular value decomposition. (we will see this relation shortly...)

<u>Singular Value Decomposition</u>

SVD is a straight forward extension of ideas of evd to the case of rectangular matrices.

In case of square matrices, we said "we are free to choose any basis (other than the std. basis), ~~and~~ but give the basis in which the matrix looks diagonal" (in the sense that the transformation represented by the original matrix in the std. basis is same as that represented by a diagonal matrix in the chosen basis).

Now a rectangular matrix $M_{m \times n}$ represents a lin. transformation from $\mathbb{R}^n \to \mathbb{R}^m$. So it is appropriate to say "you are free to choose any basis for $\mathbb{R}^n$ & any basis for $\mathbb{R}^m$, but choose them such that you are able to represent the transformation represented by M in the std. basis as a diagonal matrix with the new chosen bases for $\mathbb{R}^n$ & $\mathbb{R}^m$."

*[Realize that in the case $n=m$, according to this we are free to choose two bases one in source, one in target to make the matrix diagonal. But according to this setting i.e. evd setting we are insisting on same basis on both sides!]

So one can say that svd is an easier problem than evd!

Also for the svd setting we insist on a orthonormal bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ to be chosen rather than "any" basis. It turns out that for <u>any rectangular</u> matrix such bases can be chosen!

Here is formal discussion:

Recall that a matrix is nothing but a way of representing ~~the~~ a lin. transformation:

$$M = \begin{array}{c|c} & \begin{array}{cccc} T(e_1) & T(e_2) & \cdots & T(e_n) \end{array} \\ \hline \begin{array}{c} \\ \end{array} & \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & & \vdots \\ a_{m1} & & \cdots & a_{mn} \end{bmatrix}_{m\times n} \begin{array}{c} \to e_1 \\ \to e_2 \\ \vdots \\ \to e_m \end{array} \end{array}$$

$e_1, e_2, \dots e_n$ ~~are~~ form std. basis of $\mathbb{R}^n$

$e_1, e_2, \dots e_m$ form std. basis of $\mathbb{R}^m$

Say ~~Now~~ we ~~want to~~ choose new orthonormal basis of $\mathbb{R}^n$ as $\{v_1, v_2, \dots, v_n\}$

" $\mathbb{R}^m$ as $\{u_1, u_2, \dots, u_m\}$

then ~~the~~ the transformation wrt. bases will be diagonal only if

T is same

$$\begin{array}{cccc} T(v_1) & T(v_2) & \cdots & T(v_n) \end{array}$$
$$\begin{bmatrix} \cdot & & & \\ & \cdot & 0 & \\ & 0 & \cdot & \\ & & & \end{bmatrix} \begin{array}{c} \to u_1 \\ \to u_2 \\ \vdots \\ \to u_m \end{array}$$

only if there are entries on the principal diagonal are non-zero.

in other words I must choose bases such that:

$T(v_1) = Mv_1 = \sigma_1 u_1$

$T(v_2) = Mv_2 = \sigma_2 u_2$

$\vdots$

soon

~~Now~~ Now if M is of rank r then there can be only $v_1, v_2, \dots v_r$ where $Mv_i \neq 0$ $\therefore$ we must consist ~~of~~ on the following conditions:

$$Mv_1 = \sigma_1 u_1$$
$$Mv_2 = \sigma_2 u_2$$
$$\vdots$$
$$Mv_r = \sigma_r u_r$$

& $\{v_1, v_2, \dots, v_r\}$ span the row space of M

it is easy to see $u_1, u_2, \dots, u_r$ must span column space of M!

(Note that these eqns. are analogous to the e.v.p. equations!)

Now, for any orthonormal basis of $N(M)$ $(\dim(N(M)) = n - r)$

say $\{v_{r+1}, v_{r+2}, \dots, v_n\}$ we have $Mv_i = 0$.

So they do not effect the diagonal nature of the new trans. matrix

In summary,

Suppose we choose a orthonormal basis for row space of M i.e. $\{v_1, v_2, \dots, v_r\}$

& suppose we are able to choose orthonormal basis column space i.e. $\{u_1, u_2, \dots u_r\}$

such that $M v_i = \sigma_i u_i$ —① then we are done i.e.

we will be able to diagonalize the matrix using the following bases for $\mathbb{R}^n$ & $\mathbb{R}^m$:

$$\{\underbrace{v_1, v_2, \dots, v_r}_{\substack{\text{Specific orthonormal basis} \\ \text{of row space of } M \text{ such that} \\ \text{① holds.}}}, \underbrace{v_{r+1}, v_{r+2}, \dots, v_n}_{\substack{\text{any orthonormal} \\ \text{basis for the nullspace} \\ \text{of } M}}\}$$

&

$$\{\underbrace{u_1, u_2, \dots, u_r}_{\substack{\text{Specific orthonormal basis} \\ \text{of column space of } M \text{ such that} \\ \text{① holds}}}, \underbrace{u_{r+1}, u_{r+2}, \dots, u_m}_{\substack{\text{any orthonormal} \\ \text{basis for the left null space} \\ \text{of } M.}}\}$$

In a minute we will show that bases like this can be chosen.
But now we will see what if they exist:

We can summarise the equations in ① in this matrix equation:

$$M_{m\times n} V_{n\times n} = U_{m\times m} \Sigma_{m\times n} \longrightarrow \text{required diagonal matrix}$$

V: i^{th} column is v_i (orthonormal matrix)

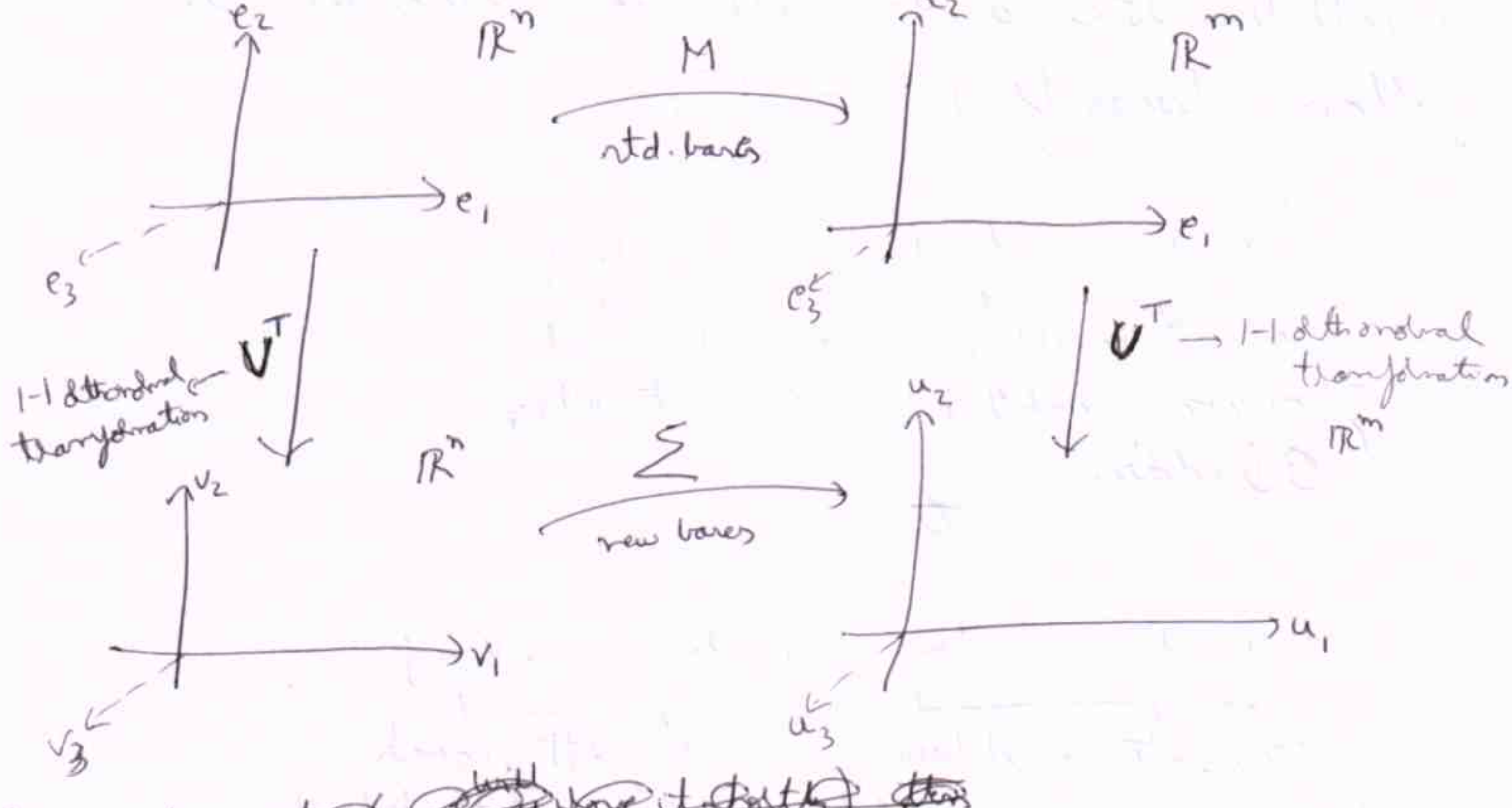
U: i^{th} column is u_i (orthonormal matrix)

$$\Sigma = \begin{bmatrix} \begin{bmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ 0 & & \sigma_n \end{bmatrix}_{n\times n} & 0 \\ 0 & 0 \end{bmatrix}_{m\times n}$$

$$\Rightarrow \boxed{M = U\Sigma V^T}$$

This is known as the singular value decomposition of a rectangular matrix M. (note analogy with expression of e.v.d.)

Here is another geometric picture



~~If svd exists (we will prove it shortly), then~~
~~$M = U\Sigma V^T \Rightarrow$ ...~~

We can also write the svd equation in a more compact form:

$$M_{m\times n} = U_{m\times m}\, \Sigma_{m\times n}\, V^T_{n\times n} \rightarrow \text{original form}$$

we know except for the left upper $r\times r$ submatrix everything is zeros

Let us call this submatrix $r\times r \rightarrow$ as $\bar{\Sigma}_{r\times r}$

U, V are orthonormal matrices

$$M = \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix} \begin{bmatrix} \bar{\Sigma}_{r\times r} & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix}$$

$$\boxed{M_{m\times n} = \bar{U}_{m\times r}\, \bar{\Sigma}_{r\times r}\, \bar{V}^T_{r\times n}} \rightarrow \text{this is the "compact form"}$$

$$\bar{U}_{m\times r} = \begin{bmatrix} u_1 & u_2, & \ldots & u_r \end{bmatrix}$$

$$\bar{V}_{n\times r} = \begin{bmatrix} v_1 & v_2, & \ldots & v_r \end{bmatrix}$$

~~Not~~ Note the $\bar{U}$ & $\bar{V}$ are not orthonormal matrices!

It is easy to see that $\bar{V}^T\bar{V} = I$ but $\bar{V}\bar{V}^T \neq I$

$|||^{ly}$ $\bar{U}^T\bar{U} = I$ but $\bar{U}\bar{U}^T \neq I$

Also, once we have the compact form, we have:

$$M^TM = \bar{V}\bar{\Sigma}\bar{U}^T\bar{U}\bar{\Sigma}\bar{V}^T = \bar{V}\bar{\Sigma}^2\bar{V}^T \rightarrow \text{nothing but e.v.d. of } M^TM$$

$$\& \; MM^T = \bar{U}\bar{\Sigma}\bar{V}^T\bar{V}\bar{\Sigma}\bar{U}^T = \bar{U}\bar{\Sigma}^2\bar{U}^T \rightarrow \text{nothing but e.v.d. of } MM^T$$

Following conclusions can be drawn (assuming s.v.d. of M exists):

(i) eig. val. of M^TM = eig. val. of MM^T

(ii) $\{v_1, v_2, \ldots, v_r\}$ are nothing but the eigenvectors of M^TM & $\sigma_i = \sqrt{\lambda_i}$ are the sq. roots of e. val. of M^TM

(iii) $\{u_1, u_2, \ldots u_r\}$ are nothing but the eigenvectors of MM^T & $\sigma_i = \sqrt{\lambda_i}$ are the sq. roots of e. val. of MM^T. [In (i) (ii) (iii) eig. val means non-zero eig. val.]

This shows if at all svd exists then the (i), (ii), (iii) hold. Now we will prove the converse (i.e. existence of SVD):

<u>Proof</u> Proof is again by induction similar in spirit to that of Spectral theorem. Here is the sketch:

Suppose we want to solve

$$\max_{\|u\|=1,\ \|v\|=1} u^T M v$$

→ $u^T M v$ must be simultaneously maximized wrt both u & v which is difficult to do

this is same as $\max_{\|u\|=1,\ \|v\|=1} \|Mv\| \cos(Mv, u)$ → this is easy because once the v such that $\|Mv\|$ is max. is chosen, then u can be aligned with Mv to get $\cos\theta = 1$

*$\left[\cos(Mv, u) = \dfrac{u^T M v}{\|u\| \, \|Mv\|} = \dfrac{u^T M v}{\|Mv\|} \right]$

we already know that v that max. $\|Mv\|$ is the "first eigenvector" of $M^T M$! i.e. optimal $v = v_1$ → eigenvector corres. to max.

Also $\|Mv_1\| = \sqrt{\lambda_1}$ → eigen value of $M^T M$

($\lambda_1 \neq 0$ otherwise $M^T M$ is zero matrix!)

Now we can choose u to be some $\alpha M v_1$ (same dir. as of Mv) ($\alpha \geq 0$)

but $\|u\| = 1 \Rightarrow \alpha \|Mv_1\| = 1 \Rightarrow \alpha = \dfrac{1}{\sqrt{\lambda_1}}$

$$\therefore \quad u = \frac{1}{\sqrt{\lambda_1}} M v_1$$

$$\Rightarrow \boxed{M v_1 = \sqrt{\lambda_1}\, u}$$

→ already looks like svd problem

Again the above optimization problem is also equivalent to

$$\max_{\|v\|=1,\ \|u\|=1} \|M^T u\| \cos(M^T u, v)$$

again the optimal is achieved when $u = u_1 \rightarrow$ the "first" eigenvector i.e. corresponding to max. eig. val of MM^T

also $\|M^T u_1\| = \sqrt{\lambda_1}$

& v must be chosen in same dir as $M^T u$, & with length 1

this again says $v = \frac{1}{\sqrt{\lambda_1}} M^T u_1$ i.e. $M^T u_1 = \sqrt{\lambda_1}\, v_1$

$\therefore$ The optimal of the opt. prob is characterized by:

$u \longrightarrow u_1 \rightarrow$ "first" eigenvector of MM^T

$v \longrightarrow v_1 \rightarrow$ " " " " of $M^T M$

& $M v_1 = \sqrt{\lambda_1}\, u_1$

$\sqrt{\lambda_1}$ → nothing but σ_1 singular value of M.

Now once this is done we can look at the orthogonal complement of v_1 i.e. all v such that $v \perp v_1$

we can show that $Mv \perp u_1$

$$\begin{aligned} & Mv \perp u_1 \\ \Longleftrightarrow\ & u_1^T M v = 0 \\ \Longleftrightarrow\ & v^T M^T u_1 = 0 \\ \Longleftrightarrow\ & v^T \sqrt{\lambda_1}\, v_1 = 0 \\ \Longleftrightarrow\ & v^T v_1 = 0 \end{aligned}$$

Now we can consider the orthogonal complement of $v_1 \underset{\text{subspace}}{\subseteq} R(M)$

" " " of $u_1 \underset{\text{subspace}}{\subseteq} C(M)$

|||ly

Here again we can choose

$$M v_2 = \sqrt{\lambda_2}\, u_2 \quad ; \; M^T u_2 = \sqrt{\lambda_2}\, v_2$$

so on until we have chosen $\left.\begin{array}{l}\\ \\ \end{array}\right\}$ r basis vectors for the row as well as column space satisfying the svd equations.

The rest of basis vectors can be any orthonormal basis of ~~corresponding~~ null spaces of M & M^T respectively.

Hence existence of SVD for any rectangular matrix is proved.

$\longrightarrow$

In summary, any matrix

$M_{m\times n}$ ~~can~~ be decomposed as

$$\underbrace{U_{m\times m}}_{\text{orthonormal}} \; \underbrace{\Sigma_{m\times n}}_{\text{diagonal}} \; \underbrace{V^T_{n\times n}}_{\text{orthonormal}}$$

~~Suppose we collect col.~~ The immediate advantage is

1. All answers about four fundamental subspaces of a matrix are exposed in SVD.

For eg rank, ~~bo~~ orthogonal basis of all ~~sub~~ 4 fundamental subspaces etc are all readable from this decomposition!

$\underline{\underline{eg}}$ Suppose $\Sigma = \begin{bmatrix} 0 & & & & \\ & 2 & & 0 & \\ 0 & & 3 & & \\ & & & 0 & \\ & & & & 0 \\ & & & & 0 \end{bmatrix}_{6\times 5}$

then immediate conclusions are:

(i) rank $= 2$, $\dim N(M) = 5-2 = 3$, $\dim N(M^T) = 6-2 = 4$.

(ii) Orthonormal basis of row space is v_2, v_3 $\rightarrow$ 2, 3 columns in V

" " " is u_2, u_3 $\rightarrow$ 2, 3 columns in U

orthonormal basis of null space is $\{v_1, v_4, v_5\}$

" " left " " is $\{u_1, u_4, u_5, u_6\}$

Other applications of SVD: (assume from now $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_n$) without loss of generality

② Can be used to ~~define~~ compute the generalized (pseudo) inverse we defined in one of the previous classes.

(we will not prove here)

It can be shown that $M^{\dagger} = V \Sigma^{\dagger} U^T$

generalized inverse of M — generalized inverse of Σ

If $\Sigma = \begin{bmatrix} \bar{\Sigma} & 0 \\ 0 & 0 \end{bmatrix}$ (as we have written earlier), then $\Sigma^{\dagger} = \begin{bmatrix} \bar{\Sigma}^{-1} & 0 \\ 0 & 0 \end{bmatrix}$

($\bar{\Sigma}$: $r\times r$; 0: $r \times (n-r)$, $m-r\times r$, $m-r\times n-r$; in $\Sigma^{\dagger}$: $r\times r$, $r\times m-r$, $n-r\times r$, $(n-r)\times(m-r)$)

$\bar{\Sigma}^{-1}$ always exists because singular values by definition are non-zero & positive.

In the special cases,

M is full column rank i.e. $n = r$ (say $m > n$)

then we know left inverse $(M^TM)^{-1}M^T$ exists. We will see that $M^{\dagger}$ will match with $(M^TM)^{-1}M^T$

~~In this~~ $M^+ = V\Sigma^+U^T$ for full column rank case:

Now, $M^+M = V\Sigma^+U^T U\Sigma V^T$

$= V\Sigma^+\Sigma V^T \longrightarrow \begin{bmatrix}\bar{\Sigma}^{-1} & 0\end{bmatrix}\begin{bmatrix}\bar{\Sigma} \\ 0\end{bmatrix} = I + 0 = I$

$= VV^T = I$

but $MM^+ = U\Sigma V^T V\Sigma^+U^T$

$= U\Sigma\Sigma^+U^T$

$= U\begin{bmatrix}\bar{\Sigma} \\ 0\end{bmatrix}\begin{bmatrix}\bar{\Sigma}^{-1} & 0\end{bmatrix}U^T$

$= U\begin{bmatrix} I_{n\times n} & O_{n\times m-n} \\ O_{m-n\times n} & O_{m-n\times m-n}\end{bmatrix}U^T \neq I_{m\times m}$

Now $M(M^TM)^{-1}M^T = U\Sigma V^T(V\bar{\Sigma}^2V^T)^{-1}V\Sigma^TU^T$

$= U\Sigma V^TV\bar{\Sigma}^{-2}V^TV\Sigma^TU^T$

$= U\left(\Sigma\bar{\Sigma}^{-2}\Sigma^T\right)U^T$

$= U\left(\begin{bmatrix}\bar{\Sigma} \\ 0\end{bmatrix}\bar{\Sigma}^{-2}\begin{bmatrix}\bar{\Sigma} & 0\end{bmatrix}\right)U^T$

$= U\left(\begin{bmatrix}\bar{\Sigma} \\ 0\end{bmatrix}\begin{bmatrix}\bar{\Sigma}^{-1} & 0\end{bmatrix}\right)U^T$

$= U\begin{bmatrix} I_{n\times n} & O_{n\times m-n} \\ O_{m-n\times n} & O_{m-n\times m-n}\end{bmatrix}U^T$

$\therefore$ In this case it matches with M^+. Similarly in case M is of full column rank, M^+ will match with $M^T(MM^T)^{-1}$.

However once we have the SVD, we can compute generalized inverses even for cases $m, n > r$.

Another main application is the following:

Suppose we have a matrix $M_{m \times n}$ and we need an approximation of this matrix $\bar{M}$ such that rank $\bar{M} <$ rank M

Say rank of $M = r$ we want to find $\bar{M}$ having rank $\bar{r} < r$ such that "$\bar{M}$ is close to M".

In other words we want to find $\bar{M}$ with rank $\bar{r}$ such that $\|M - \bar{M}\|_F$ is minimized

$\longrightarrow$ Frobenius norm we studied earlier gives distance between M & $\bar{M}$.

It turns out that $\bar{M}$ is exactly the matrix

$$U \tilde{\Sigma} V^T \quad \text{if } M = U \Sigma V^T$$

where
$$\tilde{\Sigma} = \begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & 0 & \\ & & \ddots & & & \\ & & & \sigma_{\bar{r}} & & \\ & 0 & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix}$$

Pick only the top $\bar{r}$ singular values. Put the rest of them zero. We are assuming the singular values are arranged in descending order of mag value.

eg given $M = U \Sigma V^T$ say

what is the "best rank 1" approximation of M? it will be

$$U \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ & 0 & & 0 \\ & & \ddots & \\ 0 & & & 0 \end{bmatrix} V^T \quad \text{i.e. } \underline{u_1 \sigma_1 v_1^T}$$

This notion of approximating a matrix with another matrix of lower rank is used in many applications (involving PCA)

~~To read in~~ <u>Further reading</u>:

(i) Prof. Abhiram's tutorial "Some uses of spectral methods"
http://www.cse.iitb.ac.in/~ranade/miscdocs/svd.pdf
(This is an excellent article please read it)

(ii) Jonathon Shlens "A tutorial on PCA" principal component analysis
www.snl.salk.edu/~shlens/pub/notes/pca.pdf

(iii) The following video lectures in Prof. Strang's MIT course:
1, 3, 6, 9, 10, 14, 17, 21, 22, 25, 27, 29, 30, 33 <u>This is a MUST</u>

In Endsem there might be a question from any of these lectures.

(iv) In pdf Chp 1, 2, 3, 5, 6, 7, 9 (but only relevant topics)

(v) In Prof. Strang's book: Chp 1, 2, 3, 5, 6.

THANK YOU