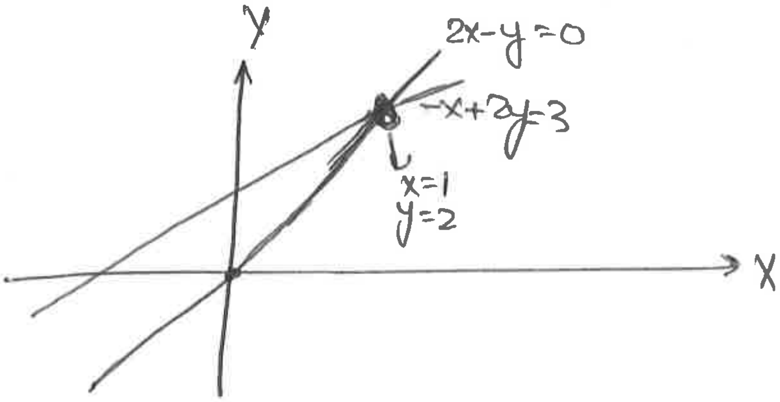
Given $2x - y = 0$
$-x + 2y = 3$

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$A \quad x = b$$

Row Picture



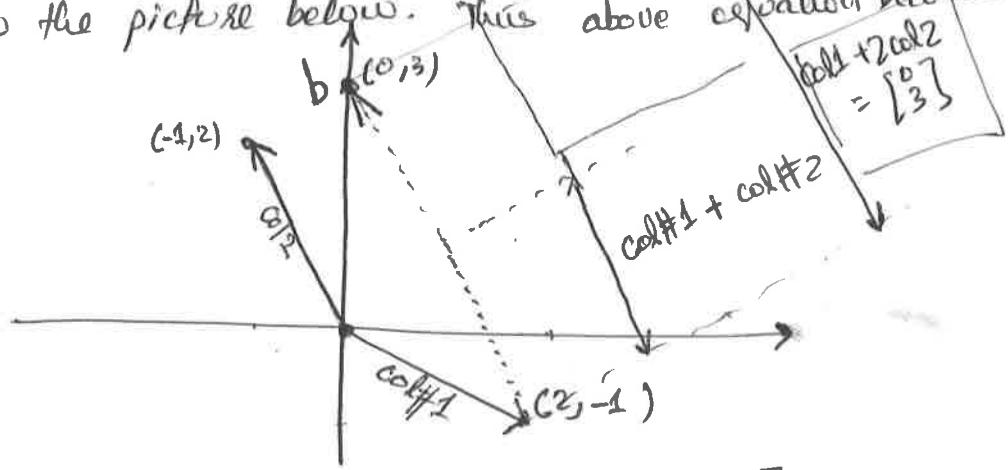
Column Picture

$$x \begin{bmatrix} 2 \\ -1 \end{bmatrix} + y \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

This is called linear combination of column

Lets draw it:

As we know that solution is (1,2), we can use it to draw the picture below. This above equation becomes $1\begin{bmatrix} 2 \\ -1 \end{bmatrix} + 2\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$



i.e we combine col 1 & col 2 to get "b" which is $\begin{bmatrix} 0 \\ 3 \end{bmatrix}$.

Now 3x3 example:

$$\begin{aligned} 2x - y \quad &= 0 \\ -x + 2y - z &= -1 \\ -3y + 4z &= 4 \end{aligned}$$

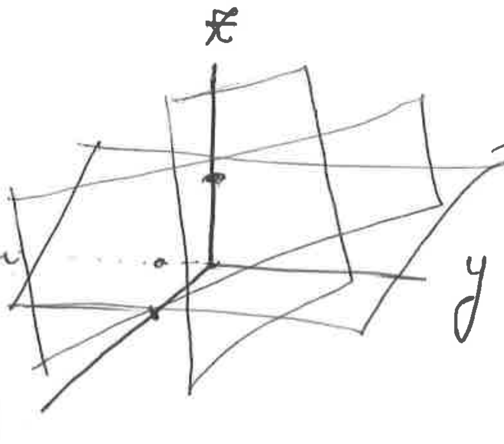
we can understand this system of equation using ~~col~~ row way (row picture) or column picture (important way).

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -3 & 4 \end{bmatrix} \qquad b = \begin{bmatrix} 0 \\ -1 \\ 4 \end{bmatrix}$$

These 3D equations are plane

Row Picture

Thus row picture is getting difficult with increased dimension

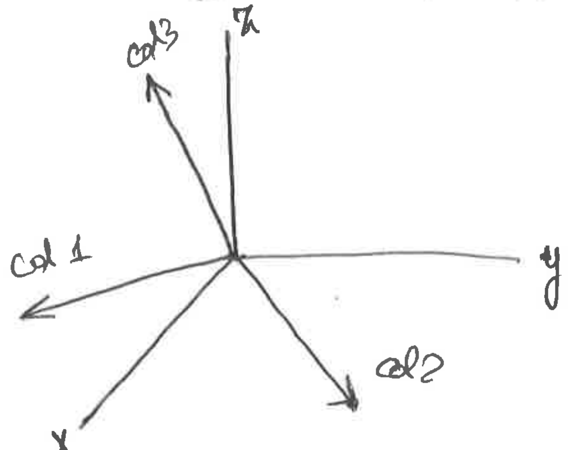


picture of eq① ② & ③

Now we will go to column picture

$$x \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} + y \begin{bmatrix} -1 \\ +2 \\ -3 \end{bmatrix} + z \begin{bmatrix} 0 \\ -1 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 4 \end{bmatrix}$$

(linear combination of three vectors)



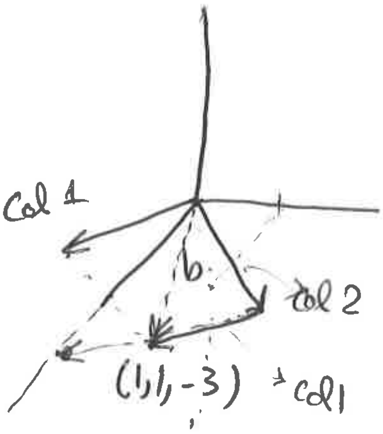
Now the above equation is asking to combine three vectors (col1, col2 & col 3) so that their linear combination produces $\begin{bmatrix} 0 \\ -1 \\ 4 \end{bmatrix}$

By looking at equation in above (column picture), we can see that solution is $x=0$ $y=0$ $z=1$

Now changing the column picture to get new b:

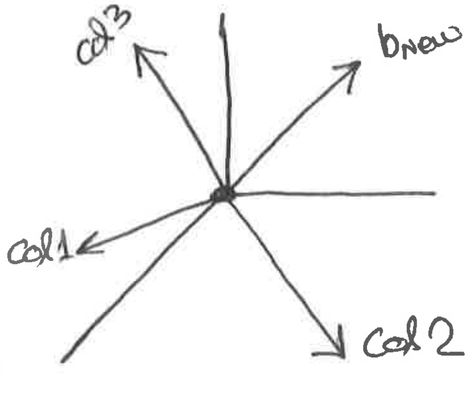
$$x\begin{bmatrix}2\\-1\\0\end{bmatrix}+y\begin{bmatrix}-1\\2\\-3\end{bmatrix}+z\begin{bmatrix}0\\-1\\4\end{bmatrix}=\begin{bmatrix}1\\1\\-3\end{bmatrix}$$

The solution is $x=1$ $y=1$ $z=0$



Can I solve $Ax=b$ for every b?
OR in linear combination words we can pose this question:
Do the linear combination of the columns fill 3-D space?
For this matrix A $\begin{bmatrix}2&-1&0\\-1&2&-1\\0&-3&4\end{bmatrix}$, the answer is yes.

When will the answer be "No"? When A is singular. Lets see diagram?



Lets say we want to reach bnew by a linear combination of col 1, col 2, col 3. When can we not reach bnew?

Ans: If col 1, col 2, col 3 lies in same plane, then their combination will lie in the same plane. At that situation, A will be singular. For instance if col 3 is the sum of col 2 & col 1, then all of them lie in same plane. Hence the only "b" in $Ax=b$ which I can get will be in the same plane! Thus we can solve it for those b's which are in the same plane & all those b's outside that plane would be unreachable. That is singular case.

For 2x2 system:

$$Ax = b$$

$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = ??$$ How to multiply them

① Inner Product form:

$$\begin{bmatrix} 2(1)+5(2) \\ 1(1)+3(2) \end{bmatrix} = \begin{bmatrix} 12 \\ 7 \end{bmatrix}$$

② Linear combination of columns:

$$1\begin{bmatrix} 2 \\ 1 \end{bmatrix} + 2\begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 12 \\ 7 \end{bmatrix}$$ Hence $Ax = b$ is a combination of columns of "A".

System of equations:

$$x + 2y + z = 2$$
$$3x + 8y + z = 12$$
$$4y + z = 2$$

To solve: $Ax = b$

Elimination — Success / FAILURE

Backsubstitution

Elimination

Matrix multiplication

$$\begin{matrix} \boxed{1} & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{matrix} \xrightarrow[\substack{(2,1) \\ \text{is wiped out}}]{R_2 - 3R_1} \begin{matrix} \boxed{1} & 2 & 1 \\ 0 & \boxed{2} & -2 \\ 0 & 4 & 1 \end{matrix}$$

1st pivot

A

2nd pivot

$$\begin{matrix} \boxed{1} & 2 & 1 \\ 0 & \boxed{2} & -2 \\ 0 & 4 & 1 \end{matrix} \xrightarrow[\text{be wiped out}]{(3,2)\text{ should}} \begin{matrix} \boxed{1} & 2 & 1 \\ 0 & \boxed{2} & -2 \\ 0 & 0 & \boxed{5} \end{matrix}$$

U

3rd pivot

* U = upper triangle matrix
* pivots can't be zero
* Determinants = product of pivot in U.

How could this have failed? If 3rd pivot was -4 initially and we would be unable to exchange rows to make pivot elements non-zero. Then the process of elimination gives "failure".

BACK substitution:

$$\underbrace{\begin{matrix} \boxed{1} & 2 & 1 & 2 \\ 3 & 8 & 1 & 12 \\ 0 & 4 & 1 & 2 \\ & A & & b \end{matrix}}_{\text{Augmented matrix}} \xrightarrow{(3,1)} \begin{matrix} \boxed{1} & 2 & 1 & 2 \\ 0 & \boxed{2} & -2 & 6 \\ 0 & 4 & 1 & 2 \end{matrix} \xrightarrow{(3,2)} \begin{matrix} \boxed{1} & 2 & 1 & 2 \\ 0 & \boxed{2} & -2 & 6 \\ 0 & 0 & \boxed{5} & -10 \end{matrix}$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad U \qquad C$$

Now final equations which are told by U, C are

$$x + 2y + z = 2$$
$$2y - 2z = 6$$
$$5z = -10$$

This is backsubstitution As the system is triangular

Now $z = -2,\ y = 1,\ x = 2$

Matrices

$$\begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix}$$

Now remember, when we multiply matrix (A) with vector (X), we can write the product as a linear combination of columns.

for instance

$$\underset{A}{\begin{bmatrix} - & - & - \\ - & - & - \\ - & - & - \end{bmatrix}} \underset{X}{\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}} = 3 \times \text{col } 1 + 4 \times \text{col} 2 + 5 \times \text{col} 3$$

$$\boxed{\text{matrix} \times \text{column} = \text{column}}$$

Now have a look on rows ~~operator~~ multiplication with matrix.

$$\begin{bmatrix} 1 & 2 & 7 \end{bmatrix} \begin{bmatrix} - & - & - \\ - & - & - \\ - & - & - \end{bmatrix} = \left.\begin{matrix} 1 \times \text{row } 1 \\ + 2 \times \text{row } 2 \\ + \cancel{3}\,7 \times \text{row } 3 \end{matrix}\right\} \text{This is linear combination of rows.}$$

$$\boxed{\text{row} \times \text{matrix} = \text{row}}$$

Matrices Again:

$$\begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix}$$ How do we get to $\begin{matrix} 1 & 2 & 1 \\ 0 & \cancel{8}\,2 & -2 \\ 0 & 4 & 1 \end{matrix}$?

Subtract 3x row 1 from row 2

$$\begin{bmatrix} & E_{21} \text{ matrix} & \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{bmatrix}$$

How to write E_{21}. Use linear combination of rows.

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{E_{21}} \begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{bmatrix}$$

STEP 2 : subtract $2 \times R_2$ from R_3

$$E_{32} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 0 & 5 \end{bmatrix}$$

(E_{32} fixing (3,2) pos)

OR we can write :

$$E_{32}(E_{21}A) = U$$

OR removing parenthesis using associative law of multiplication

$(E_{32}E_{21})A = U$ Hence we can get U in single shot.

Another type of elementary matric. is the matric that exchanges two ~~matrix~~ rows. It is called permutation matrix.

Permutation:

Exchange rows 1 and 2

$$\underset{P}{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} c & d \\ a & b \end{bmatrix}$$

If we want to exchange columns of a matrix ?

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \underset{P}{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}} = \begin{bmatrix} b & a \\ d & c \end{bmatrix}$$

In short, to do column operation we multiply on right & to do row operation we multiply on left!

In matrix multiplication $A \cdot B \neq B \cdot A$

The better way to solve $E_{32}(E_{21}A) = U$ is to think how can we go from U to A.

Inverses:

$$\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

we got this matrix E_{21} by doing something. How to inverse this matrix. ~~by adding R2~~ Simply by nullifying change.

$$\underset{E^{-1}}{\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}\underset{E}{\begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}} = \underset{I}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}$$

Short - INTRO for LEAST SQUARES:

i) Projection of vector onto subspace.

ii) Projection in high dimension

iii) Connecting Projection with least square (Joining to Normal equation)

iv) Visualize ~~primal~~ C(A)

v) Visualize ~~dual~~ $N(A^T)$

vi) Mention Applications.

vii) Types: WEIGHTED LEAST SQUARES (~~proje~~ oblique projection)
REGULARIZED LEAST SQUARES

Questions

Q1. why do we need LS?

Disclaimer

→ Linear least squares

→ Only it intuition.

→ All variants left

EXTRA DERIVING Normal equation using

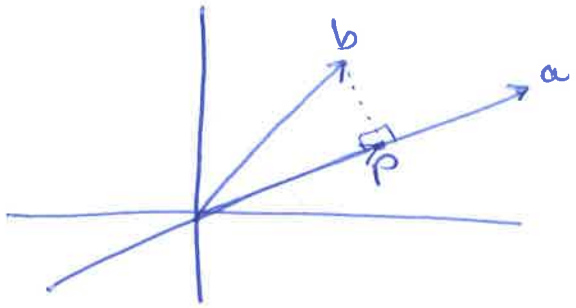
i) CALCULUS (Pg 307, Boyd)

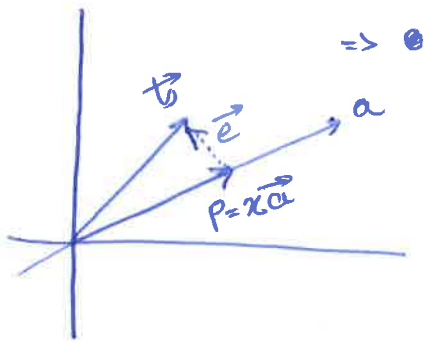
ii) Least square solution of linear equation (least-norm problem) (Pg ~~6~~304, Boyd)

iii) $A^TA x^* = A^T b$ (Normal equations of least squares problem Pg 458, BOYD).

* MENTION RANK Deficiency (Mathworks leastsquares)

Projections onto subspace





$\Rightarrow$ $\vec{b} = \vec{p} + \vec{e}$

* "b" projected on $\vec{a}$. $\vec{p}$ is an approximation of $\vec{b}$

$$p = (\vec{a}) \frac{a^T b}{a^T a}$$

OR

$$\boxed{\vec{p} = \left(\frac{a^T b}{a^T a}\right)\vec{a}}$$

Subspace: $x\vec{a} \in$ Subspace $\vec{a}$

$x_1\vec{b} + x_2\vec{c} \in$ subspace

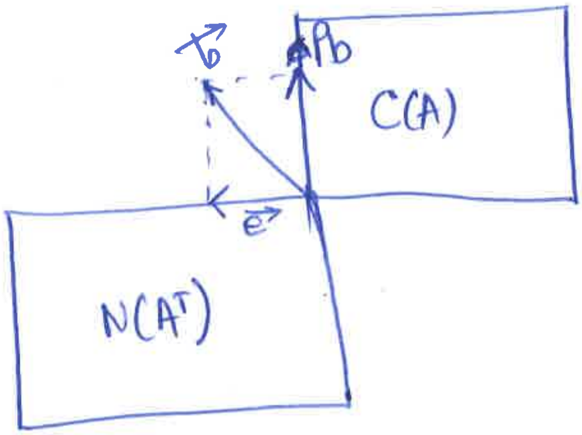
REASON

WHY Projection is under discussion

PROJECTION In high dimensions.

$$P = A(A^TA)^{-1}A^T$$ → If A = square, invertible then it column span $R(A)$.

* Project any vector onto $R(A) \Rightarrow P\cdot(\text{vector})$
* Project b onto $R(A)$: $Pb \in R(A)$



MATRIX Projecting $\vec{b}$ onto $N(A^T)$ is $I-P$:

$$e = b - p$$
$$e = \underbrace{(I-P)}_{\text{projection matrix (Naturally projection matrix)}}\, b$$

it has all properties of pro[...]

* Could put simple example of linear regression to show application of least squares.

$$A^T A \hat{x} = A^T b \quad \text{(solves least squares)}$$

jethion

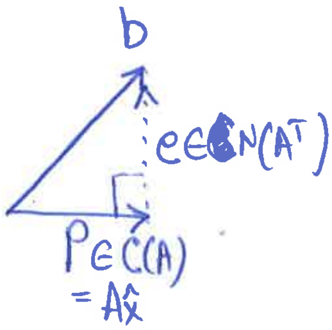
REACHING NORMAL EQUATION:

$$Ax = b \quad \left(\begin{array}{l}\text{unsolvable} \\ b/c\ b \notin R(A)\end{array}\right)$$

FUNDAMENTAL THEOREM of LA:

$$C(A) \perp N(A^T)$$

i) PROJECT "b" onto $C(A)$ as p



ii) $b = p + e$, with $p \in C(A)$ and $e \in N(A^T)$

$$b = \underbrace{A\hat{x}}_{} + e$$

~~LETS CONSIDER 2D~~

Find the right combination of columns of "A" such that "$e = b - p$" $\underbrace{\text{is} \perp}$ to ~~plane~~ basis of $C(A)$

Key to projection

$\Rightarrow \quad e = b - A\hat{x}$

$e \perp \vec{a_1}$ $\qquad$ $e \perp \vec{a_2}$

$a_1^T(b - A\hat{x}) = 0$ $\quad$ OR $\quad$ $a_2^T(b - A\hat{x}) = 0$

$$\begin{bmatrix} a_1^T \\ a_2^T \end{bmatrix}(b - A\hat{x}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$A^T(b - A\hat{x}) = 0$$

$$\boxed{A^TA\hat{x} = A^Tb} \quad \textit{Normal equation}$$

Application:

$\cancel{y = Ax + b}$ $\quad y = mx + c \longrightarrow \begin{cases} y_1 = mx_1 + c \\ y_2 = mx_2 + c \\ \vdots \\ y_n = mx_n + c \end{cases}$

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \quad X = \begin{bmatrix} m \\ c \end{bmatrix} \quad b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$AX = b$ (inconsistent)

$$\hat{X} = (A^T A)^{-1} A^T b$$

$$\|Ax - b\|_2^2 = \left\| \begin{bmatrix} mx_1 \\ mx_2 \\ \vdots \\ mx_n \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \right\|_2^2 = \left\| \begin{bmatrix} mx_1 - y_1 \\ mx_2 - y_2 \\ \vdots \\ mx_n - y_n \end{bmatrix} \right\|_2^2$$

$$= \left(\sqrt{(mx_1 - y_1)^2 + \cdots + (mx_n - y_n)^2} \right)^2$$

$$= (mx_1 - y_1)^2 + \cdots + (mx_n - y_n)^2$$

$$= \sum_{i=1}^{n} (mx_i - y_i)^2$$

L_2-Norm:

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

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

PCA

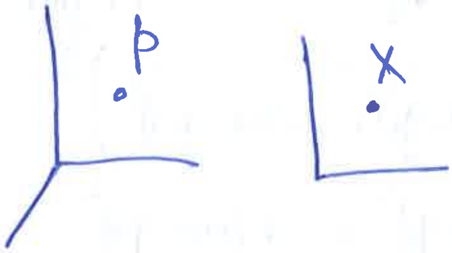
$$A \longrightarrow W = \begin{pmatrix} | & | \\ e_1 & e_2 \\ | & | \end{pmatrix}$$

$$P = W \underbrace{(W^T W)^{-1}}_{I} W^T$$

$$\boxed{P = WW^T}$$

$$X = W^T p$$

$$p = WX$$



$$p = WW^T b$$
$$= Pb$$

$$P = \left| \begin{matrix} \circ & \circ & \circ \\ \circ & \circ & \circ \\ \circ & \circ & \circ \end{matrix} \right| \left| \begin{matrix} \circ \\ \circ \\ \circ \end{matrix} \right|$$

$\{b_i\}$

$$\min_{W} \sum_{i=1}^{N} (\underbrace{b_i - P_i}_{e_i})^2$$

$$= \sum_{i=1}^{N} (b_i - WW^T b_i)^2$$

Maximizing ~~Minimizing~~ Variance is equivalent to minimizing squared reconstruction error.

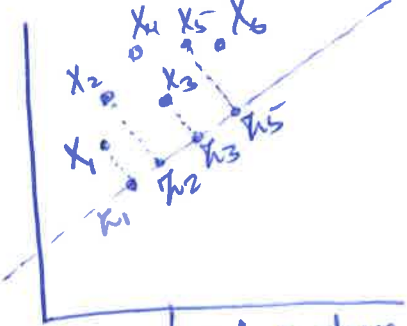
$X = x_1, x_2 \cdots x_n$

L_K = projection matrix = [orthonormal columns]

$\tilde{x}_i = x_i^K = V_K^T x_i$

No. of. dimension after PCA = K

Basis ~~Eigenvectors~~ for this plane in $V_K = \begin{bmatrix} | & | & & | \\ e_1 & e_2 & \cdots & e_K \\ | & | & & | \end{bmatrix}$



PCA on

$x_1, x_2 \cdots x_n \in R^P$

$q < P$

$q \to$ reduced dimension

$P \to$ original dimension

$f(\lambda) = \mu + V_q \lambda$ (Reconstruction of R^q into R^P)

$f(\lambda) \to R^P$, $V_q \to P \times q$, $\lambda \to$ ~~output of PCA~~ Reduced dim data points (R^q)

Reconstruction error e_n $\begin{cases} = x_n - f(\lambda_n) \\ = x_n - (\mu + V_q \lambda_n) \\ = x_n - \mu - V_q \lambda_n \end{cases}$

As $\lambda_n = V_q^T x_n$

$\Rightarrow e_n = x_n - \mu - V_q(V_q^T x_n)$

$e_n = x_n - \mu - V_q V_q^T x_n$

$\Rightarrow \min_{V_q} \|e_n\|^2$

$\min \|x_n - \mu - V_q V_q^T x_n\|^2$

$\min_{V_q} \sum_{n=1}^{N} \|x_n - V_q V_q^T x_n\|^2 \longrightarrow$ PCA

$W \to$ plane on which data is projected

Projection on $C(W)$

$P_W = W(W^T W)^{-1} W^T$

$\boxed{P_W = W W^T}$

~~Let $P \in R^3$ & $x \in R^2$~~

Project vector "b" onto low dimensional space given by W:

$P_b = \underline{W(W^T W)^{-1} W^T}\, b$

$P_{b_i} = W W^T b_i$ in high dimension

$W^T b_i \to$ same point in low-dim

$x_n \in R^P$

$V_q \in R^{P \times q}$

PCA

Project data ("b") from high dimensional space to low dimensional space such that the error incurred by reconstructing the data in high dimensional space is minimized

$\min_W \sum_{i=1}^{N} (b_i - W W^T b_i)^2$

Topic 3: Optimization (+ Duality)

*) Discrete in R^n include linear programming, quadratic programming with constraints.

* Continuous where the unknown is function $U(x,y)$. It is "calculus of Variations".

In continuous optimization, we want to find derivative $=0$ for unknown function $U(x,y)$ [must see].

Start with Discrete in R^n

Primal (we state first)

LEAST SQUARES: $\min \|AU-b\|^2$ $\quad A = m\times n$, $m>n$

if $A = m\times m$, then $\min\|AU-b\|^2$ is solving $AU=b$.

$m\times n$ → # of unknowns

# of measurement or # of equations (size of b)

"CALCULUS" will lead us to best "u".

"Normal Equation" $\quad A^TA\,\hat{U} = A^Tb$ (solves least squares)

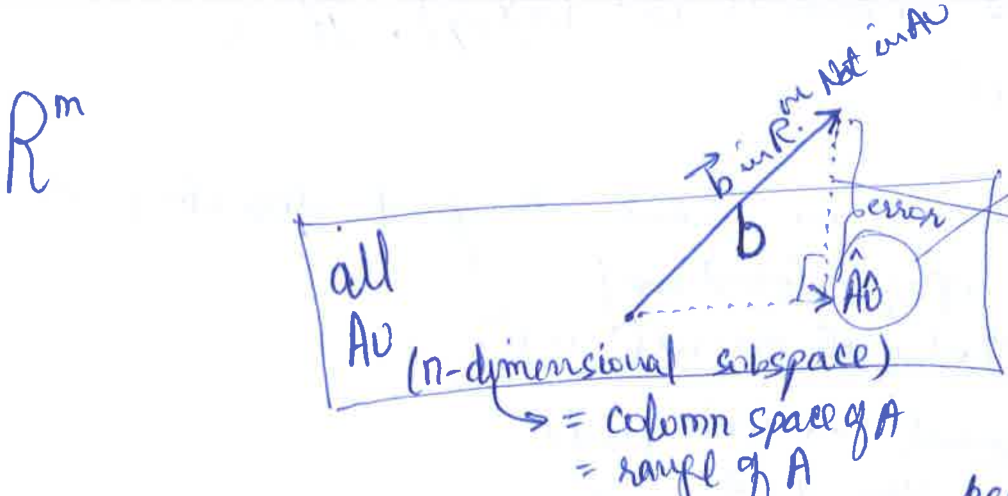
Name comes from statistics

i) symmetric

ii) PD (means A has rank="n" or columns of A are independent)

~~Lets Assume A has~~

We want optimization with Duality. In optimization, very often problems in optimization are two problem. The two problems in least square is.

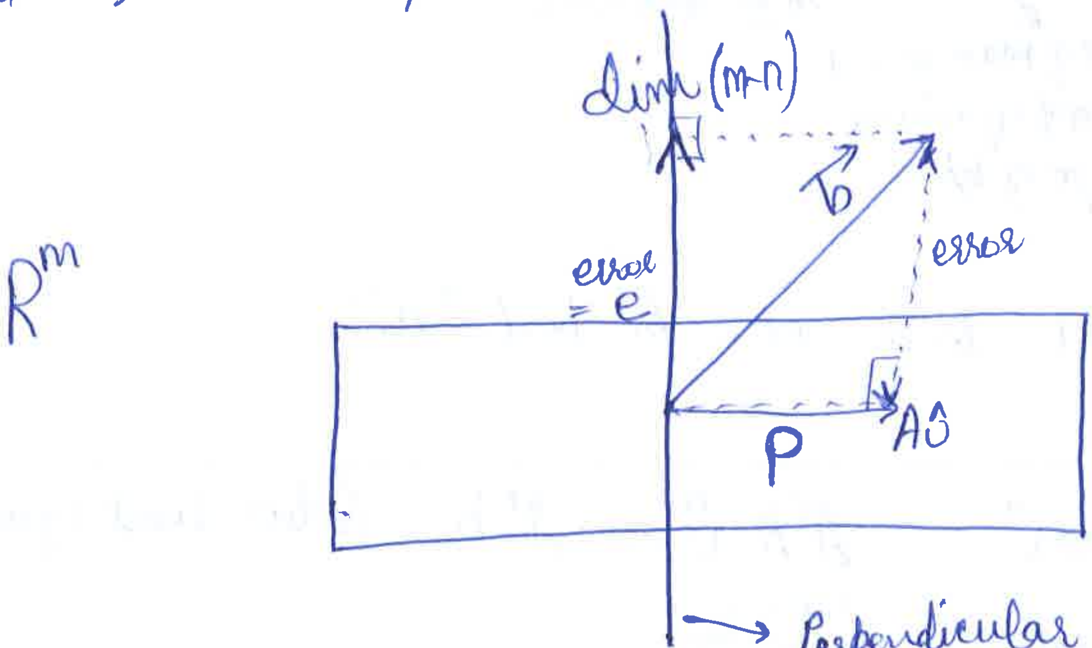


Best Av closest to $\vec{b}$ in $C(A)$.

We want to minimize this error. We can see pythagoras here.

The closest Av ($A\hat{v}$) to $\vec{b}$ is line ("error") perpendicular to the plane $C(A)$. Thus $A\hat{v}$ is projection or orthogonal projection of $\vec{b}$ on $C(A)$.

Lets see dual problem:



Perpendicular subspace to $A\hat{v}$. All vectors perpendicular to $C(A)$. Its dim is $m-n$.

To get perpendicular subspace to $C(A)$ will give us view of dual problem of least squares.
Dual problem is actually find $\vec{e}$ (projection of $\vec{b}$) on plane $\perp$ to $C(A)$.

Thus $$b = p + e$$
p = projection on column space
e = projection on ~~plane~~ subspace $\perp$ to $C(A)$

Let y be subspace $\perp$ to $C(A)$.
OR y in perp subspace
OR y is perp to column A

$$\Rightarrow \underbrace{\begin{bmatrix} \text{— col } 1^T \text{ —} \\ \text{— col } 2^T \text{ —} \\ \vdots \\ \text{— col } n^T \text{ —} \end{bmatrix}}_{A^T} \begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

(Dual problem)

$\Rightarrow \quad A^T y = 0 \quad$ (e is one among "y")

i.e y is "nullspace of A^T". $N(A^T)$ is $\perp$ $C(A)$. [fundamental theorem of Linear Algebra].

Thus Dual problem : $\boxed{\min \|b-y\|^2 \text{ with } A^T y = 0}$ ← Sol = e

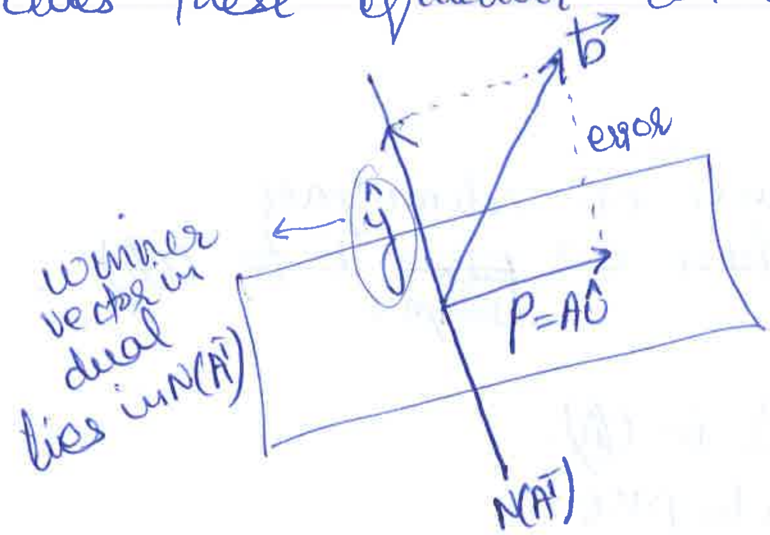
$\boxed{\text{Primal Problem: } \min \|Au-b\|^2}$ ← sol = $\hat{u}$
$P = A\hat{u}$

What does Dual / Primal equation says.

i) Primal is unconstrained with n unknown

ii) Dual is constrained with $m-n$ unknown

How does these equation connect ?



$$\Rightarrow \left.\begin{aligned} \hat{y} + A\hat{u} &= b \\ A^T\hat{y} &= 0 \end{aligned}\right\} \quad \underbrace{\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} \hat{y} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} \overset{\text{Data}}{b} \\ 0 \end{bmatrix}}_{\text{Connection of Primal \& dual at same time.}}$$

No. 1 method to solve Linear Programming was simplex. We can't use Above method to solve LP because LP contains inequality constraints.

A new method called Primal - Dual will solve Linear programming using above equation.

Lets look at

$$\begin{bmatrix} C^{-1} & A \\ A^T & 0 \end{bmatrix}$$

rectangular

Also PD

This is also a matrix which solves many applications. Compare with $\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix}$

In weighted Least square, Au is weighted by some matrix often covariance matrix. Lets assume its C^{-1}

Lets see.

$$\left[\begin{array}{ccc|cc} \boxed{1} & 0 & 0 & 2 & 5 \\ 0 & \boxed{1} & 0 & 3 & 6 \\ 0 & 0 & \boxed{1} & 4 & 7 \\ \hline 2 & 3 & 4 & 0 & 0 \\ 5 & 6 & 7 & 0 & 0 \end{array}\right]$$

i) It is invertible

ii) 3 Positive pivots. 2 Negative Pivots (see below)

$\Big\downarrow$ elimination

$$\begin{bmatrix} I & A^T \\ 0 & \boxed{-A^TA} \end{bmatrix}$$

-29

or $\begin{bmatrix} 2 & 3 & 4 \\ 5 & 6 & 7 \end{bmatrix}\begin{bmatrix} 2 & 5 \\ 3 & 6 \\ 4 & 7 \end{bmatrix} = \begin{bmatrix} 29 & \end{bmatrix}$

As we have negative pivots while elimination step, therefore we can say that its a saddle point problem. Also we can not use stochastic gradient descent here. In order to use it we should do reduction here.

$$\begin{bmatrix} I & A^T \\ \hline 0 & -A^TA \end{bmatrix}\begin{bmatrix} y \\ u \end{bmatrix} = \begin{bmatrix} \hat{b} \\ 0 \end{bmatrix}$$

$\longrightarrow$ reduction so as to use stochastic gradient descent.

In $\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix}\begin{bmatrix} \hat{y} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}$ $m+n$ equations

y's $\quad$ u's

Normal equation should come from above pair. Natural way to solve Above system is elimination.

Block system #1

$$\boxed{\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix}\begin{bmatrix} \hat{y} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}}$$

$$\text{(i)} \Rightarrow A^T \hat{y} + A^T A \hat{u} = A^T b \quad (\times A^T)$$

$$\text{(ii)} \Rightarrow A^T \hat{y} = 0$$

As $A^T y = 0$

$$\Rightarrow \text{(i)} \Rightarrow \boxed{A^T A \hat{u} = A^T b}$$

Normal equation

$$\begin{bmatrix} I & A \\ 0 & -A^T A \end{bmatrix}\begin{bmatrix} \hat{y} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} b \\ -A^T b \end{bmatrix}$$

* Saddle-Point System
* Optimality equation
* Kuhn-Tucker equations

optimization world

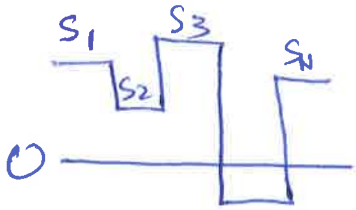
Why it is Saddle-Point?

Lets look at $\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix}$.

Weighted Least Squares

Question

optimize heat distribution (close to given $u_0(x)$). Choose heat sources $S_1, S_2 \cdots S_N$



$$\rightarrow \frac{d^2u}{dx^2} = S(x) \rightarrow u(x) \quad \text{to minimize} \int_0^1 (u-u_0)^2 \, du$$

Picture

We want heat distribution like

$$\underbrace{\frac{d^2u}{dx^2}}_{\text{output}} = \underbrace{S(x)}_{\text{Input}}$$

S is input, u is output & we want to $\min \int_0^1 (u-u_0)^2 \, du$ or we want u to match u_0.

ORDINARY LEAST SQUARES

$$e + Au = b$$
$$A^T e = 0$$

$$S = \begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix}$$

WEIGHTED LEAST SQUARES

$$A^T A \hat{u} = A^T b$$

$$\min \|WAu - Wb\|^2 \rightarrow (WA)^T (WA) \hat{u}_w = (WA)^T Wb$$

$$\boxed{A^T C (b - A\hat{u}) = 0} \leftarrow A^T C A \hat{u}_w = A^T C b$$

$\hat{u}_w$ Now depends on weight

$C = W^T W$ (PSD)

$\boxed{\text{Now measurements are weighted}}$

Normal equation for weighted least squares:

$$A^T C A \hat{u}_w = A^T C b$$

$$\Rightarrow A^T C A \hat{u}_w - A^T C b = 0$$

$$A^T C (b - A\hat{u}) = 0$$

Ordinary L.S.

Primal Dual Equation

$$e + Au = b$$

$$A^T e = 0$$

Advantage of two equations is that we get $\hat{u}$ as well as total error "e".

$$S = \begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix}$$

Weighted least SQ

Primal Dual EQUATION

$$e + Au_w = b$$

$$A^T C e = 0$$

$$S = \begin{bmatrix} I & A \\ A^T C & 0 \end{bmatrix}$$

I don't want this b/c it destroys symmetry.

We will introduce new variable

$w = Ce \Rightarrow e = C^{-1} w$

* Saddle Point System
* KKT system
* Primal dual pair

New Equations

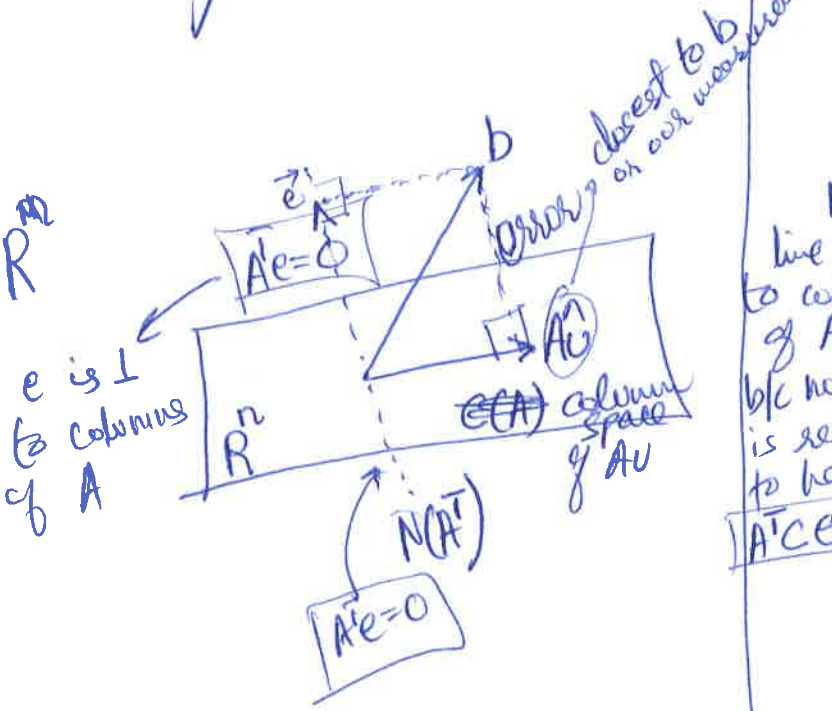
$$C^{-1} w + Au_w = b$$

$$A^T w = 0$$

$$\Rightarrow S = \begin{bmatrix} C^{-1} & A \\ A^T & 0 \end{bmatrix}$$

S = (saddle point matrix)

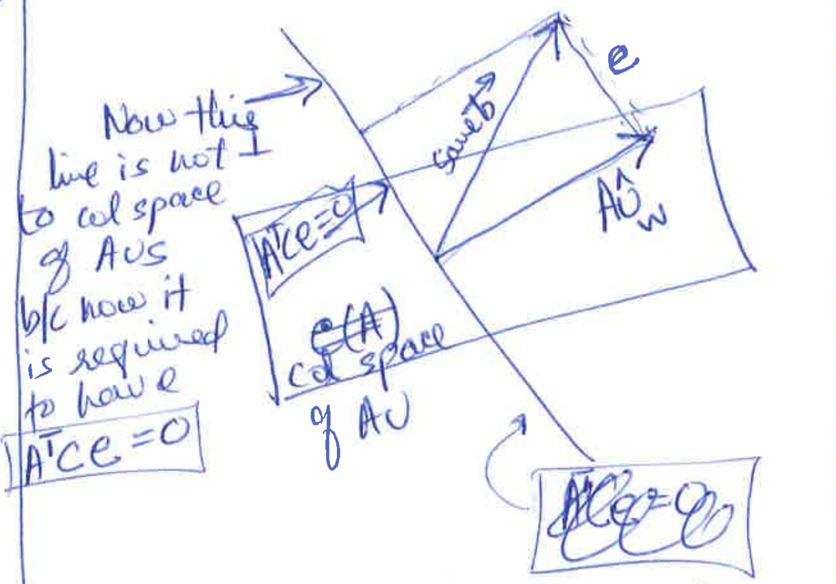
Ordinary Least Squares



Here $X^Ty = 0$

Weighted Least Squares

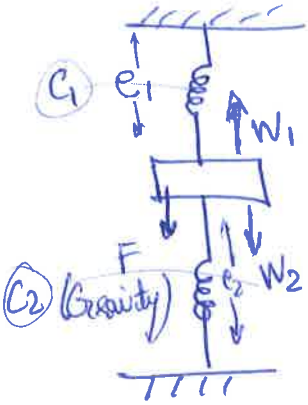
Change of geometry when W is introduced



Now $\vec{b}$ is projected on $A^TCe = 0$. Now we dont have right triangle. Its //gm where $\vec{b}$ has a ~~block~~ oblique projection on $A^TCe = 0$.

Here $X^TCy = 0$ i.e it is C-orthogonal.

Mechanics



$W_1 = W_2 + F$ (equilibrium)

$\Rightarrow w_1 - w_2 = F$

Now Minimize $E(w) = E_1(w_1) + E_2(w_2)$ } minimize energy in spring

s.t. $w_1 - w_2 = f$

LAGRANGE

$$L(w,u) = \left[E_1(w_1) + E_2(w_2)\right] - u(w_1 - w_2 - f)$$

Lagrange multiplier

what sign?

Now $\dfrac{\partial L}{\partial w_1} = 0 = \dfrac{dE_1}{dw_1} - u = \dfrac{w_1}{c_1} - u = 0$

$\dfrac{\partial L}{\partial w_2} = 0 = \dfrac{dE_2}{dw_2} + u = \dfrac{w_2}{c_2} - u = 0$

$\dfrac{\partial L}{\partial u} = 0 \Rightarrow w_1 - w_2 - f = 0$ (constraint equation) similarly to $A^T w = 0$ as in Least squares.

As $E_1(w_1) = \frac{1}{2} c_1 e_1^2$ $E_2(w_2) = \frac{1}{2} C_2 e_2^2$

spring constant   stretch

Hooks law says that $w = ce \Rightarrow e = w/c$

9

$\Rightarrow \quad E(w) = \frac{1}{2}\frac{w_1^2}{c_1} + \frac{1}{2}\frac{w_2^2}{c_2}$

$\Rightarrow \quad L(w,u) = \frac{1}{2}\frac{w_1^2}{c_1} + \frac{1}{2}\frac{w_2^2}{c_2} - (w_1 - w_2 - f)$

$\Rightarrow \quad \frac{w_1}{c_1} - u = 0$

$\frac{w_2}{c_2} + u = 0$

$\boxed{w_1 - w_2 = f}$

we dont want this b/c it destroys symmetric

symmetric part

$$\begin{bmatrix} 1/c_1 & 0 & -1 \\ 0 & 1/c_2 & +1 \\ +1 & -1 & 0 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \\ u \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ f \end{bmatrix}$$

Anti-Symmetric Part

SADDLE POINT form

Could be

$$\begin{bmatrix} 1/c_1 & 0 & -1 \\ 0 & 1/c_2 & +1 \\ -1 & +1 & 0 \end{bmatrix} \; i$$

if $w_2 - w_1 = f$

LAGRANGE Variable interpretation

MEANING of u

It is the amount of mass comes down
or it is displacement

Turns out $\quad u = \frac{dE_{min}}{df} = \text{sensitivity}$ (change in minimum energy wrt change in input)