- **Linear Transformation**
- **Projections and least squares approximations**

Transformation Matrix *A*

- **Transformation:** $x \rightarrow Ax$ ($R^n \rightarrow R^m$ if A is m by n)
- **Stretch:** $A=cI$ **expands or contracts the vectors.**

Ex. $A=\begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix}$

- **Rotation: turns the space around the origin.**

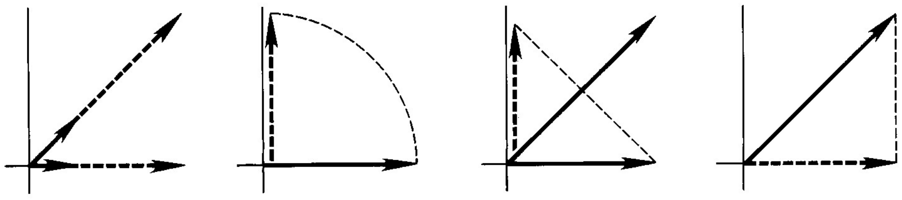
Ex. $A=\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ **turns all vectors through 90°.**

- **Reflection: transforms every vector into its mirror image.**

Ex. $A=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ **the mirror is the 45° line** ($x=y$)

- **Projection: project a space onto a lower-dimensional subspace. (projection matrix must not be invertible)**

Ex. $A=\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ **transforms vector** (x, y) **in the plane to the nearest point** (x, 0) **on the horizontal axis.**



Linear Transformation

- **What a matrix transformation cannot possibly do:**

(i) It is impossible to move the origin since $A0=0$

(ii) If A **transforms** x **to** x', **then** cx **can be only transformed to** cx' **since** $A(cx)=c(Ax)$

(iii) If A **transforms** x **to** x' **and** y **to** y', **then their sum** $x+y$ **can only be transformed to** $x'+y'$ **since** $A(x+y)=Ax+Ay$

- **Any matrix transformation follows these three rules.**
- **Three rules can be combined into:**

For all numbers c and d and all vectors x **and** y, **matrxi multiplication satisfies the rule of *linearity*:**

$$A(cx+dy)=c(Ax)+d(Ay)$$

Any transformation that meets this requirement is a *linear transformation*

- **Every matrix is a linear transformation. Question: *Does every linear transformation lead to a matrix*? That is, can we express any transformation in matrix form?**

Linear Transformation Example - Polynomials

- **Polynomial space P_n: $p=a_0+a_1t+\ldots+a_nt^n$ with space dimension = n+1**
- **Example: *differentiation* operation: $A=d/dt$**

$$Ap=\frac{d}{dt}(a_0+a_1t+\cdots+a_nt^n)=a_1+\cdots+na_nt^{n-1}$$

Nullspace: one-dimensional space of constant polynomial

Column space: n-dimensional space P_{n-1}

- **Example: *integration* from 0 to t**

$$Ap=\int_0^t(a_0+\cdots+a_nt^n)dt=a_0t+\cdots+\frac{a_n}{n+1}t^{n+1}$$

nullspace: none (except for the zero vector as always)

column space: P_{n+1} without constant term

- **Example: Multiplication by a fixed polynomial, say 2+3t**

$$Ap=(2+3t)(a_0+\cdots+a_nt^n)=2a_0+\cdots+3a_nt^{n+1}$$

It transforms P_n to P_{n+1} and no nullspace except for p=0

- **Nonlinear transformations: square ($Ap=p^2$), adding 1 ($Ap=p+1$) or**

keeping the positive terms ($A(t-t^2)=t$),…

Basis and Linear Transformation

- **Let $x_1, x_2,\ldots,x_n$ be the basis, then any other vector in the space is a combination of the vectors in the basis:**

$$\text{if } x=c_1x_1+\cdots+c_nx_n \text{ then } Ax=c_1(Ax_1)+\cdots+c_n(Ax_n)$$

- **Crucial property:**

Once Ax_i's are determined, Ax of any x is known

- <u>**A vector $x=(c_1, c_2,\ldots, c_n)$ is actually defined to be the linear combination of the basis $x=c_1x_1+\cdots+c_nx_n$ and $c_1, c_2,\ldots, c_n$ are coefficients of the linear combination.**</u>

- **Example: differentiation for the polynomials of degree 3. The natural choice of basis vector: $p_1=1$, $p_2=t$, $p_3=t^2$, $p_4=t^3$**

All we need to know is how the transformation performs on the basis: $Ap_1=0$, $Ap_2=1$, $Ap_3=2t$, $Ap_4=3t^2$

Let $p_1=(1, 0, 0, 0)$, $p_2=(0, 1, 0, 0)$, $p_3=(0, 0, 1, 0)$, $p_4=(0, 0, 0, 1)$

Then, $A=\begin{bmatrix}0&1&0&0\\0&0&2&0\\0&0&0&3\\0&0&0&0\end{bmatrix}$ **found by the basis transformations.**

- **This matrix can be then used to differentiate any polynomials of degree 3: say, $p=2+t-t^2-t^3$**

$$\frac{dp}{dt}=Ap\rightarrow\begin{bmatrix}0&1&0&0\\0&0&2&0\\0&0&0&3\\0&0&0&0\end{bmatrix}\begin{bmatrix}2\\1\\-1\\-1\end{bmatrix}=\begin{bmatrix}1\\-2\\-3\\0\end{bmatrix}\rightarrow 1-2t-3t^3$$

Linear Transformation in Matrix Form

- **Question: how to find matrix *A*?**
- ***Suppose the vectors $x_1, \ldots, x_n$ are a basis for the space V and $y_1, \ldots, y_m$ are a basis for W. Then each linear transformation from V to W is represented by a matrix A:***

  Apply transformation to the j^{th} basis vector of V, x_j, and then express the transformation result as a combination of the W basis :$Ax_j=A(0x_1+0x_2+\ldots+1x_j+\ldots+0x_n)=a_{1j}y_1+a_{2j}y_2+\ldots+a_{mj}y_m$

  i.e. $A\begin{bmatrix}0\\ \vdots\\ 1\\ \vdots\end{bmatrix}\begin{matrix}1^{st}\\ \vdots\\ j^{th}\\ \vdots\end{matrix}=\begin{bmatrix}a_{1j}\\ a_{2j}\\ \vdots\\ a_{mj}\end{bmatrix}.$

  Then, $a_{1j}, a_{2j}, \cdots, a_{mj}$ ***form the j^{th} column of A.*** **(Why?)**

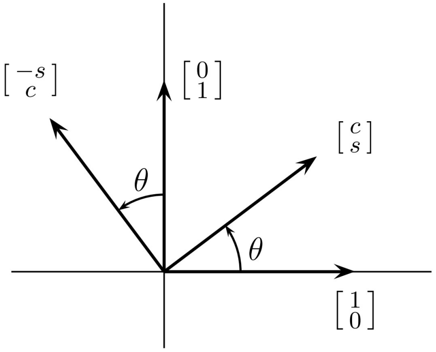
Example: find the transformation matrix for integration on a polynomial of degree 3. Basis: *x*: 1, *t*, *t*², *t*³ ; *y*: 1, *t*, *t*², *t*³, *t*⁴

$$\int_0^t 1dt = t \text{ or } Ax_1 = y_2, \cdots, \int_0^t t^3 dt = \frac{1}{4}t^4 \text{ or } Ax_4 = y_5 \quad \Rightarrow$$

$$A_{\text{int}} = \begin{bmatrix}0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 1/2 & 0 & 0\\ 0 & 0 & 1/3 & 0\\ 0 & 0 & 0 & 1/4\end{bmatrix} \quad A_{\text{diff}}A_{\text{int}} = I \text{ BUT } A_{\text{int}}A_{\text{diff}} \neq I \text{ !!}$$

Rotation Matrix *Q*

- **Rotation through an angle θ:**



$$Q_\theta = \begin{bmatrix}c & -s\\ s & c\end{bmatrix} \quad c = \cos\theta, \; s = \sin\theta$$

- **Does the inverse of Q_θ equal $Q_{-\theta}$?**

$$Q_\theta Q_{-\theta} = \begin{bmatrix}c & -s\\ s & c\end{bmatrix}\begin{bmatrix}c & s\\ -s & c\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

- **Does the product of Q_θ and Q_φ equal $Q_{\theta+\varphi}$?**

$$Q_\theta Q_\varphi = \begin{bmatrix}\cos\theta\cos\varphi - \sin\theta\sin\varphi & -\\ \sin\theta\cos\varphi + \cos\theta\sin\varphi & -\end{bmatrix} = \begin{bmatrix}\cos(\theta+\varphi) & -\sin(\theta+\varphi)\\ \sin(\theta+\varphi) & \cos(\theta+\varphi)\end{bmatrix} = Q_{\theta+\varphi}$$

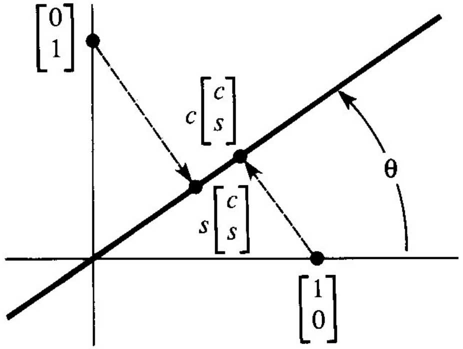
$\Rightarrow$ ***the product of the matrices corresponds to the product of the transformations***

$\boxed{A: V \to W \text{ and } B: U \to V \Rightarrow AB: U \to W}$

- **To construct *A* and *B*, we need bases for *V* & *W* and *U* & *V*. If the basis for *V* is kept the same, then the product matrix goes directly from the basis in *U* to basis in *W*.**

Projection Matrix P

- **Projection of (1, 0) onto the θ-line:**



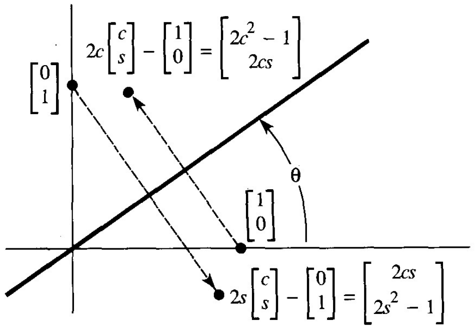
$$P = \begin{bmatrix} c^2 & cs \\ cs & s^2 \end{bmatrix}.$$

- **The transformation can not be reversed ⇒ The matrix has no inverse!!**
- **Points, like (-s, c), on the perpendicular line are projected onto the origin (zero vector) ⇒ perpendicular line is the nullspace of *P***
- **Points on the θ-line are projected to themselves ⇒ $P^2=P$**

$$P^2 = \begin{bmatrix} c^2 & cs \\ cs & s^2 \end{bmatrix}^2 = \begin{bmatrix} c^2(c^2+s^2) & cs(c^2+s^2) \\ cs(c^2+s^2) & s^2(c^2+s^2) \end{bmatrix} = P$$

Reflection Matrix H

- **Reflection of (1, 0) in the θ-line:**



$$H = \begin{bmatrix} 2c^2-1 & 2cs \\ 2cs & 2s^2-1 \end{bmatrix}$$

- **Two reflections bring back the original: $H^2=I \Rightarrow H^{-1}=H$**

  Can you guess H^{-1} simply from the matrix form? *But it's easy if you know the meaning of the what transformation the matrix represents.*
- **The sum of a vector and its mirror image equals twice the projection of the vector ⇒ $Hx+x=2Px \Rightarrow H=2P-I$**

  $\Rightarrow H^2=(2P-I)^2=4P^2-4P+I=I$ since $P^2=P$

Basis for Constructing Transformation Matrix *A*

Let the basis be (c, s) and ($-s$, c): one is on the θ-line and the other is on θ-line's perpendicular line.

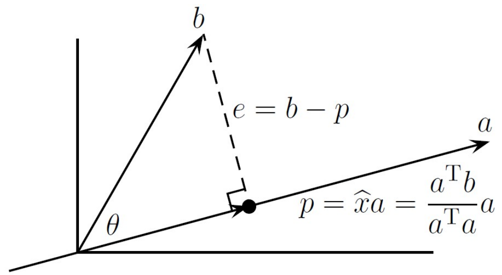
- **Projection matrix:** $P = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
- **Reflection matrix:** $H = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
- **Check $H=2P-I$**
- **Rotation matrix:** $Q = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}$

Choosing the best basis is very important!!

- **Idea: make the transformation matrix *diagonal*, as P and H here using (c, s) and ($-s$, c) as basis.**
- **By changing basis, transformation matrix A → $S^{-1}AS$ where S account for the change of basis.**
- **We come back later to this subject of choosing the best basis.**

Projection of *b* onto Vector *a*

- **The projection point p must be some multiple of a ($p = \hat{x}a$)**
- **The line from b to the closest point $p = \hat{x}a$ is perpendicular to the vector a**



$(b - \hat{x}a) \perp a$, or $a^T(b\text{-}\hat{x}a) = 0$, or $\hat{x} = \dfrac{a^T b}{a^T a}$

- **The projection of b onto the line through O and a is**

$$p = \hat{x}a = \frac{a^T b}{a^T a} a$$

Schwarz Inequality

- **Schwarz Inequality: any two vectors satisfy**

$$|a^T b| \le \|a\| \|b\|$$

Reason 1: $\|b-p\|^2 = \left\| b - \frac{a^T b}{a^T a} a \right\|^2 = \frac{(b^T b)(a^T a) - (a^T b)^2}{a^T a} \ge 0$

$\Rightarrow\ (b^T b)(a^T a) - (a^T b)^2 \ge 0\ \Rightarrow\ |a^T b| \le \|a\| \|b\|$

Reason 2: $-1 \le \cos\theta \le 1 \Rightarrow |\cos\theta| \le 1 \Rightarrow \left| \frac{a^T b}{\|a\| \|b\|} \right| \le 1\ \Rightarrow\ |a^T b| \le \|a\| \|b\|$

- **Equality holds if and only if *b* is a multiple of *a* ⇒ θ=0° or 180° and cosine=1 or –1 ⇒ *b=p***

- **Example: Project *b*=(1, 2, 3) onto the line through *a*=(1, 1, 1):** $\hat{x} = \frac{a^T b}{a^T a} = \frac{6}{3} = 2$

  The projection point *p=2a*=(2, 2, 2) $\Rightarrow \cos\theta = \frac{\|p\|}{\|b\|} = \frac{\sqrt{12}}{\sqrt{14}}$

  Schwarz inequality: $6 = |a^T b| \le \|a\| \|b\| = \sqrt{3}\sqrt{14}$

Projections of Rank One

- $p = \hat{x} a\ \Rightarrow\ p = a\hat{x} = a\frac{a^T b}{a^T a} = \frac{aa^T}{a^T a} b = Pb$

- ***Projection matrix*** $P = \frac{aa^T}{a^T a}$ **to project *b* onto *a***

Example: The matrix that projects any vector onto the line through *a*=(1, 1, 1) is

$$\boldsymbol{P} = \frac{\boldsymbol{aa^T}}{\boldsymbol{a^T a}} = \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

Two properties for *P*:

1. A symmetric matrix

2. P^2=P

- **Rank=1; column space: spanned by *a*=(1, 1, 1); left-null space is formed by vectors that satisfy $a^T b$=0.**

- ***P* remains the same if *a* is multiplied by a scalar.**

- **Projection onto the "θ-direction" line, i.e. the line through *a*=(cos θ, sin θ):**

$$P = \frac{\begin{bmatrix} c \\ s \end{bmatrix}\begin{bmatrix} c & s \end{bmatrix}}{\begin{bmatrix} c & s \end{bmatrix}\begin{bmatrix} c \\ s \end{bmatrix}} = \begin{bmatrix} c^2 & cs \\ cs & s^2 \end{bmatrix}$$

New Definition of Transpose

- **Old definition:** $(A^T)_{ij} = (A)_{ji}$
- **New definition: The inner product of Ax with y equals the inner product of x with $A^T y$: $(Ax)^T y = x^T(A^T y)$**
- **Inner product of x, transformed by A, with y equals the inner product of x with y transformed by A^T**
- **Proof for $(AB)^T = B^T A^T$:**

$$(ABx)^T y = [A(Bx)]^T y = (Bx)^T A^T y = x^T(B^T A^T y)$$

- **An important reminder: $(A^{-1})^T = (A^T)^{-1}$**

  $\Rightarrow$ transpose of the inverse = inverse of the transpose

Least Squares Approximate of Solution

- **$Ax=b$ is solvable $\Rightarrow$ b is in the column space**
- **When there are more equations than unknowns, b is hardly in the column space $\Rightarrow$ Find x such that Ax is as close as possible to b $\Rightarrow$ least square approximation**

Example:
$$2x = b_1 \\ 3x = b_2 \\ 4x = b_3$$

If (b_1, b_2, b_3) is not on the same line as (2, 3, 4) $\Rightarrow$ find x that minimizes the sum of squared errors:

$$E^2 = (2x - b_1)^2 + (3x - b_2)^2 + (4x - b_3)^2$$

Set $dE^2/dx=0$ to find $\hat{x} = \dfrac{2b_1 + 3b_2 + 4b_3}{2^2 + 3^2 + 4^2} = \dfrac{a^T b}{a^T a}$

- **Sum of squared errors for general cases:**

$$E^2 = (a_1 x - b_1)^2 + \cdots + (a_m x - b_m)^2 = \| ax - b \|^2$$

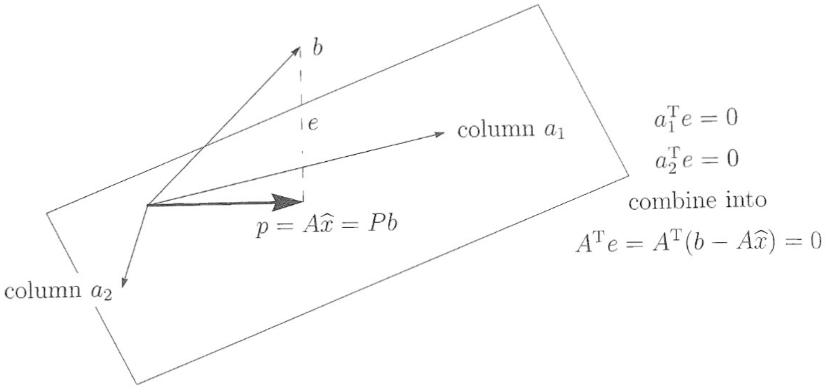
- **Set the derivative to zero $\Rightarrow$**

$$(a_1\hat{x} - b_1)a_1 + \cdots + (a_m\hat{x} - b_m)a_m = a^T(a\hat{x} - b) = 0$$

- **The least squares solution to $ax=b$ is** $\hat{x} = \dfrac{a^T b}{a^T a}$

Projection onto Column Space

- **This is a problem analogy to find an approximate solution for an unsolvable $Ax=b$ system, especially when the no. (n) of unknowns is less than the no. (m) of equations**
- **To find $\hat{x}$ that minimizes E^2 ≡ to find a point p on the column space of A that is closest to b**



⇒ **error** $e = b - A\hat{x}$ **must be perpendicular to the column space**

1. **The vectors perpendicular to the column space must lie in the *left nullspace*:** $A^T(b - A\hat{x}) = 0$ **or** $A^T A\hat{x} = A^T b$
2. **The error vector must be perpendicular to every column in A:**

$$\begin{bmatrix} a_1^T \\ \vdots \\ a_n^T \end{bmatrix} \left[b - A\hat{x} \right] = 0$$

Normal Equations, Least Squares Approximation and Projection onto a Subspace

- **The *normal equations* for an inconsistent system $Ax=b$ of m equations and n unknowns:** $\boxed{A^T A\hat{x} = A^T b}$ **which is always solvable.**
- **The least squares approximation of x exists uniquely:**

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

if $A^T A$ is invertible, i.e., the columns of A are linearly independent ($r=n$).

- **The projection of b onto the column space is:**

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

- **If b is actually in the column space of A ($b=Ax$), then:**

$$p = A(A^T A)^{-1} A^T b = A(A^T A)^{-1} A^T Ax = AIx = Ax = b$$

- **If b is perpendicular to the column space ($A^T b=0$), then:**

$$p = A(A^T A)^{-1} A^T b = 0$$

- **When A is square and invertible:**

$$p = A(A^T A)^{-1} A^T b = AA^{-1}(A^T)^{-1} A^T b = IIb = b$$

- **$(A^T A)^{-1}$ can be only taken apart when A is square and invertible.**

Projection Matrix *P*

- $p = A(A^T A)^{-1} A^T b = Pb$ $\Rightarrow$ $P = A(A^T A)^{-1} A^T$ = **Projection Matrix to project *b* onto the column space of *A***
- ***b*** **is split into *Pb* component (in the *column space*) and *b–Pb* component in the orthogonal complement (*left nullspace*)**
- ***I–P*** **is also a projection matrix that project *b* to the left null space: (*I–P*)*b*=*b–Pb***
- **Two properties for *P***

1. $P^2=P$

2. $P^T=P$ **(symmetric property)**

***Proof*:**

$$P^2 = A(A^T A)^{-1}[A^T A(A^T A)^{-1}]A^T = A(A^T A)^{-1} I A^T = P$$

$$P^T = (A^T)^T[(A^T A)^{-1}]^T A^T = (A)[(A^T A)^T]^{-1} A^T = P$$

More on Projection Matrix

- **It is known now if *P* is a projection matrix then** $P^2=P$ **and** $P^T=P$
- **Is any symmetric matrix with** $P^2=P$ **a projection matrix? YES!**

***Proof*:**

To prove *P* is a projection matrix, we have to prove that the error vector *b–Pb* is perpendicular to the component of any vector *c* projected onto that space (*Pc*):

$$(b - Pb)^T Pc = b^T (I - P)^T Pc = b^T (I - P)Pc = b^T (P - P^2)c = 0$$

$$\boxed{\text{A Projection matrix } P \Leftrightarrow P^2=P \text{ and } P^T=P}$$

Example:

If *A* is square and invertible, any vector *b* projects to itself

$\Rightarrow$ ***P*** **must be an identity matrix: *Pb*=*Ib*=*b***

How?

$$P = A(A^T A)^{-1} A^T = AA^{-1}(A^T)^{-1} A^T = I$$

Note: *I* is symmetric and $I^2=I$

A^TA and Left-Inverse of A

- $(A^TA)^T=A^TA^{TT}=A^TA \Rightarrow A^TA$ is symmetric
- A^TA has the same nullspace as A since if $Ax=0$ then $A^TAx=0$ and if $A^TAx=0$ then $x^TA^TAx=\|Ax\|^2=0=Ax$.
- Normal equation $A^T A\hat{x} = A^T b$ is always solvable but may have many solutions when A has linearly dependent columns, i.e., A *or* A^TA has non-trivial nullspace solutions.
- ***If A has linearly independent columns(r=n) with zero dimension of null space, then A^TA is an n×n square with zero dimension of null space, i.e., A^TA is square and invertible*** $\Rightarrow$ $\hat{x}$ is unique as A has full column rank
- Recall $r=n$ ($n\le m$)→ a left-inverse exist → $B_{n\times m}A_{m\times n} = I_{n\times n}$
- $r=n$ ($n\le m$) $\Leftrightarrow$ A has linearly independent columns

  $\Leftrightarrow$ ***A^TA is square, symmetric, and invertible***

  $\Leftrightarrow$ $(A^TA)^{-1}$ exists

  $\Leftrightarrow$ $(A^TA)^{-1} A^TA=I \Leftrightarrow [(A^TA)^{-1} A^T]A=I$

  $\Leftrightarrow$ $(A^TA)^{-1} A^T$ is the left-inverse of A
- **If columns of A are not linearly independent ($r<n$), the normal equation $A^T A\hat{x} = A^T b$ is still solvable with many solutions ($+x_{null}$). A new matrix, say A', has to be formed by only independent columns of A to ensure existence of $(A'^TA')^{-1}$**

Projection onto Row Space and AA^T

- Row space of A=column space of A^T

$\Rightarrow$ Projection onto row space of A

= Projection onto column space of A^T

$\Rightarrow$ Recall projection matrix to project onto A column space

$= A(A^T A)^{-1} A^T$

$\Rightarrow$ Projection matrix to project onto row space of A

= Projection matrix to project onto A^T column space

$= A^T[(A^T)^T A^T]^{-1}(A^T)^T = A^T(AA^T)^{-1}A$

if AA^T is invertible.

- ***If A has linearly independent rows(r=m), then AA^T is square, symmetric, and invertible***
- Recall $r=m$ ($m\le n$)→ a right-inverse exist → $A_{m\times n}C_{n\times m} = I_{m\times m}$
- $r=m$ ($m\le n$) $\Leftrightarrow$ A has linearly independent rows

  $\Leftrightarrow$ ***AA^T is square, symmetric, and invertible***

  $\Leftrightarrow$ $(AA^T)^{-1}$ exists

  $\Leftrightarrow$ $AA^T(AA^T)^{-1}=I \Leftrightarrow A[A^T(AA^T)^{-1}]=I$

  $\Leftrightarrow$ $A^T(AA^T)^{-1}$ is the right-inverse (pseudo-inverse) of A

AA^{T} and Pseudo-inverse

- **Recall *pseudoinverse* A^+ is to invert Ax back to the row-space part of x:**

  $A^+Ax=x_r$

- **x_r is the component of x in the row space**

  $=x_r$ is projection of x onto the row space

  $= A^T(AA^T)^{-1}Ax = A^+Ax$

  $\Rightarrow \quad A^+ = A^T(AA^T)^{-1}$

- **$A^+Ax=x_r$**

  $\Rightarrow A^+A$ is a projection matrix to project onto row space

- **$AA^+= AA^T(AA^T)^{-1} = I$**

  $\Rightarrow A^+$ is a right inverse of A *<u>when A has independent rows</u>*

- **What if AA^T is not invertible? Removing zero rows in U and deal with UU^T…**

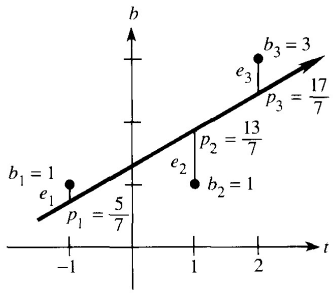
Least Squares Fitting of Data

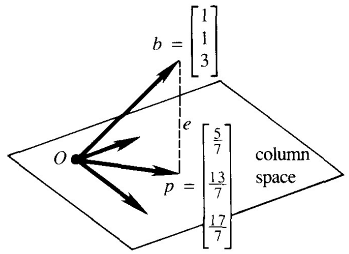
$$\begin{aligned} C + Dt_1 &= b_1 \\ C + Dt_2 &= b_2 \\ &\vdots \\ C + Dt_m &= b_m \end{aligned}$$

We obtain observations (t_1, b_1), …, (t_m, b_m) from experiments.

We would like to estimate C, and D:

$$\begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \Rightarrow \boldsymbol{Ax=b}$$





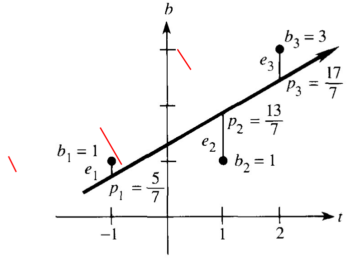
Example: b=1 at t=-1, b=1 at t=1, b=3 at t=2

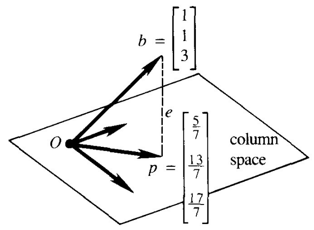
$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$$

$\Rightarrow \quad A^TA\hat{x} = A^Tb$ **is** $\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$

$\Rightarrow \quad \hat{C} = \dfrac{9}{7}$ **and** $\hat{D} = \dfrac{4}{7}$

Two Explanations of Least Squares





Right: ***b***=(1, 1, 3) is not on the plane spanned by (1, 1, 1) and (–1, 1, 2); least squares replace ***b*** by ***p*** which lies on the plane.

Left: (–1, 1), (1, 1) and (2, 3) are not one a line. Least squares replace them by points on a line: (-1, p_1) (1, p_2) (2, p_3)

- **Unable to solve *Ax*=*b*, we solve** $A\hat{x} = p$
- $e = b - A\hat{x} = b - p = (1,1,3) - (\frac{5}{7}, \frac{13}{7}, \frac{17}{7}) = (\frac{2}{7}, \frac{-6}{7}, \frac{4}{7})$ **is orthogonal to the both columns (1, 1, 1) and (-1, 1, 2)**
- **If** $b = (\frac{2}{7}, \frac{-6}{7}, \frac{4}{7}) \Rightarrow \hat{x} = 0$, **the best fitted line *y*=0 (*x*-axis)**
- **For m pairs of observations:**

$$A^T A \begin{bmatrix}\hat{C}\\ \hat{D}\end{bmatrix} = A^T b \;\Rightarrow\; \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix}\begin{bmatrix}\hat{C}\\ \hat{D}\end{bmatrix} = \begin{bmatrix}\sum b_i \\ \sum t_i b_i\end{bmatrix}$$

- $b = Ce^{-\lambda t} + De^{-\mu t}$: **to estimate *C* and *D* is a linear problem. But to estimate λ and μ is a nonlinear problem.**

Multiple Regression

Observation	Values				
	Y	X_1	X_2	…	X_p
1	y_1	x_{11}	x_{12}		x_{1p}
2	y_2	x_{21}	x_{22}		x_{2p}
⋮	⋮	⋮	⋮	…	⋮
n	y_n	x_{n1}	x_{n2}		x_{np}

- **Model:** $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + e$
- **Goal: given *n* sets of observation, estimate** $\beta_0, \beta_1, \ldots, \beta_p$

⇒ **This is a problem of *Ax*=*b*, where**

$$A = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & & x_{2p} \\ 1 & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & & x_{np} \end{bmatrix}, \quad x = \begin{bmatrix}\beta_0\\ \beta_1 \\ \vdots \\ \beta_p\end{bmatrix}, \quad b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n\end{bmatrix}$$

$$\Rightarrow \hat{\beta} = \hat{x} = (A^T A)^{-1} A^T b$$

- **The notation may confuse you but the problem is exactly the same. And this is again how mathematics can help you to solve problems that no longer can be pictured.**

Weighted Least Square

- **Sometimes observations are not trusted to the same degree because they are obtained from different accurate scales. Then, different weights can be applied to different observations when calculating the sum of squared errors:**

		Values				
Weight	**Observation**	Y	X_1	X_2	…	X_p
W_1	1	y_1	x_{11}	x_{12}		x_{1p}
W_2	2	y_2	x_{21}	x_{22}		x_{2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	…	$\vdots$
W_n	N	y_n	x_{n1}	x_{n2}		x_{np}

- **Weighted sum of squared errors:**

$$E^2 = w_1^2(\beta_0 + \beta_1 x_{11} \cdots + \beta_p x_{1p} - y_1)^2 + \cdots + w_n^2(\beta_0 + \beta_1 x_{n1} \cdots + \beta_p x_{np} - y_n)^2$$

Let *W*= $\begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & w_n \end{bmatrix} \Rightarrow E^2 = \|WAx - Wb\|^2$

That is, we would like to find a least square solution ($\hat{x}$ **) that makes** $WA\hat{x}$ **as closed as possible to** Wb**.**

$\Rightarrow$ **Least square solution to *WAx=Wb* :**

$$(A^T W^T WA)\hat{x}_W = A^T W^T Wb$$

- **In practice, choice of *W* is the most important question!**

More accurate more weight $\Rightarrow$ $w_i = 1/\sigma_{y_i}$ **(accuracy ~ 1/σ)**