LINEAR ALGEBRA

CHAPTER 4. ORTHOGONALITY

Prof. Cheolsoo Park

BCML Bio Computing & Machine Learning Lab

ORTHOGONALITY OF THE FOUR SUBSPACES

- Two vectors are orthogonal when their dot product is zero

$$v \cdot w = 0 \text{ or } v^T w = 0$$

Orthogonal vectors $\quad \boldsymbol{v}^{\mathrm{T}}\boldsymbol{w} = 0 \quad$ and $\quad \|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2 = \|\boldsymbol{v} + \boldsymbol{w}\|^2.$

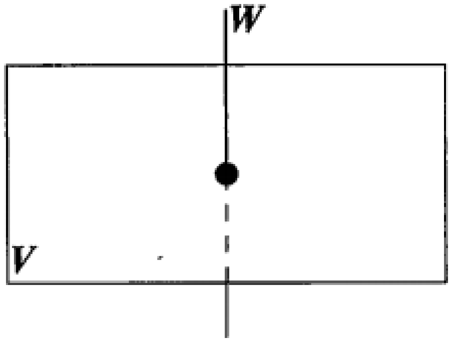
The right side is $(\boldsymbol{v} + \boldsymbol{w})^{\mathrm{T}}(\boldsymbol{v} + \boldsymbol{w})$. This equals $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}$ when $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w} = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{v} = 0$.

DEFINITION Two subspaces $\boldsymbol{V}$ and $\boldsymbol{W}$ of a vector space are ***orthogonal*** if every vector $\boldsymbol{v}$ in $\boldsymbol{V}$ is perpendicular to every vector $\boldsymbol{w}$ in $\boldsymbol{W}$:

Orthogonal subspaces $\quad \boldsymbol{v}^{\mathrm{T}}\boldsymbol{w} = 0$ ***for all*** $\boldsymbol{v}$ ***in*** $\boldsymbol{V}$ ***and all*** $\boldsymbol{w}$ ***in*** $\boldsymbol{W}$.

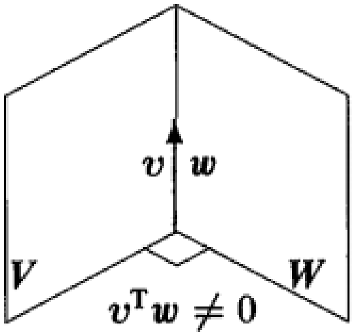
ORTHOGONALITY OF THE FOUR SUBSPACES

- Example



orthogonal line and plane

Example 2 Two walls look perpendicular but they are not orthogonal subspaces! The meeting line is in both V and W—and this line is not perpendicular to itself. Two planes (dimensions 2 and 2 in $\mathbf{R}^3$) cannot be orthogonal subspaces.



non-orthogonal planes

* Then which subspaces are orthogonal??
Go to next slide.

ORTHOGONALITY OF THE FOUR SUBSPACES

Every vector $\boldsymbol{x}$ in the nullspace is perpendicular to every row of A, because $A\boldsymbol{x} = \mathbf{0}$. ***The nullspace $N(A)$ and the row space $C(A^{\mathrm{T}})$ are orthogonal subspaces of $\mathbf{R}^n$.***

To see why $\boldsymbol{x}$ is perpendicular to the rows, look at $A\boldsymbol{x} = \mathbf{0}$. Each row multiplies $\boldsymbol{x}$:

$$A\boldsymbol{x} = \begin{bmatrix} \text{row } 1 \\ \vdots \\ \text{row } m \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \begin{array}{l} \longleftarrow \ (\text{row } 1) \cdot \boldsymbol{x} \text{ is zero} \\ \\ \longleftarrow \ (\text{row } m) \cdot \boldsymbol{x} \text{ is zero} \end{array} \qquad (1)$$

Example 3 The rows of A are perpendicular to $\boldsymbol{x} = (1, 1, -1)$ in the nullspace:

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 3 & 4 \\ 5 & 2 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{gives the dot products} \quad \begin{array}{l} 1 + 3 - 4 = 0 \\ 5 + 2 - 7 = 0 \end{array}$$

ORTHOGONALITY OF THE FOUR SUBSPACES

Every vector y in the nullspace of A^T is perpendicular to every column of A. ***The left nullspace $N(A^T)$ and the column space $C(A)$ are orthogonal in $\mathbf{R}^m$.***

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

$$A^T y = \begin{bmatrix} (\text{column } 1)^T \\ \cdots \\ (\text{column } n)^T \end{bmatrix} \begin{bmatrix} \\ y \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ \cdot \\ 0 \end{bmatrix}.$$

ORTHOGONALITY OF THE FOUR SUBSPACES

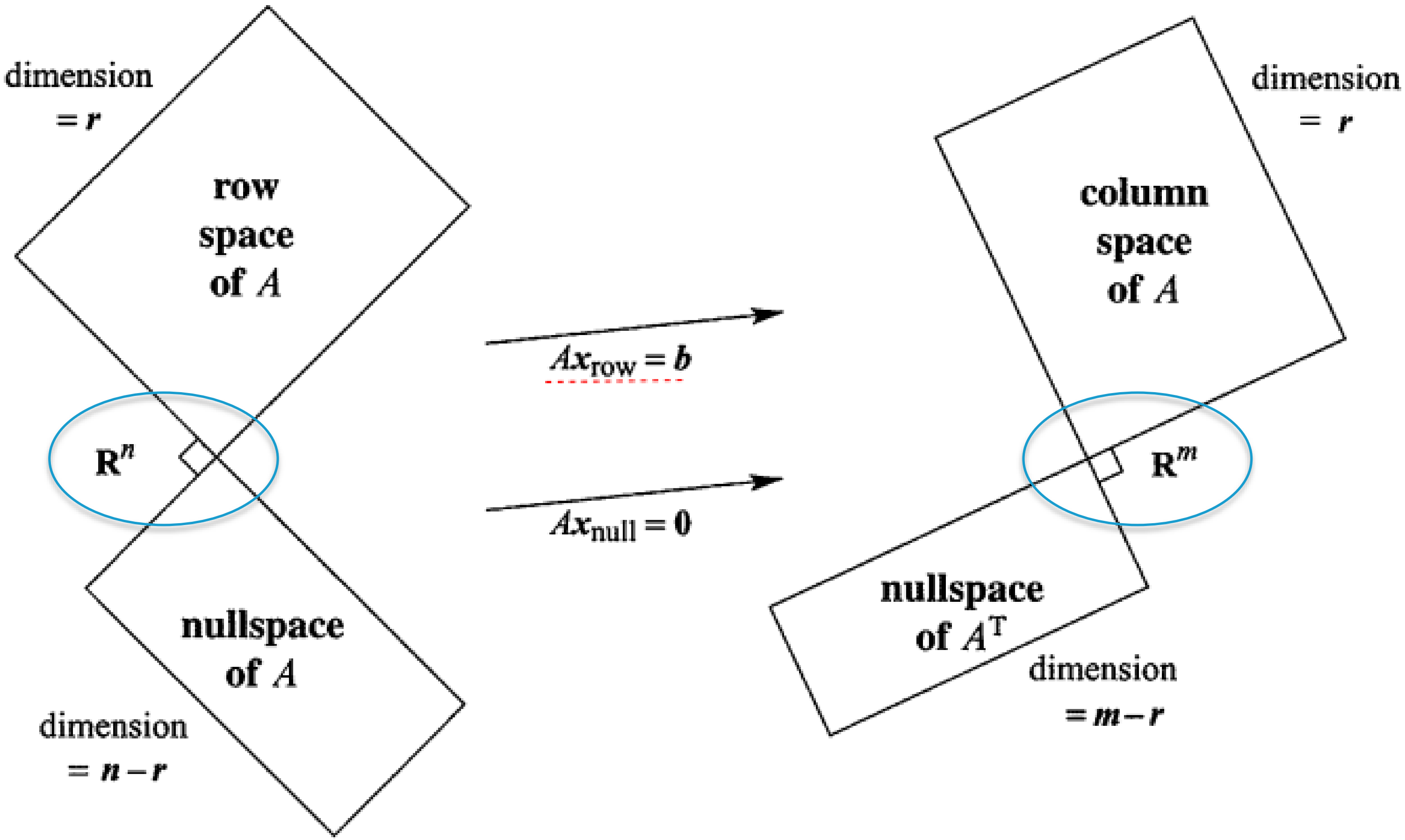


Figure 4.2: Two pairs of orthogonal subspaces. The dimensions add to n and add to m. **This is an important picture**—one pair of subspaces is in $\mathbf{R}^n$ and one pair is in $\mathbf{R}^m$.

ORTHOGONALITY OF THE FOUR SUBSPACES

- Orthogonal Complements

Not just some, but all!!!

DEFINITION The ***orthogonal complement*** of a subspace V contains ***every*** vector that is perpendicular to V. This orthogonal subspace is denoted by $V^{\perp}$ (pronounced "V perp").

By this definition, the nullspace is the orthogonal complement of the row space. ***Every*** $\boldsymbol{x}$ that is perpendicular to the rows satisfies $A\boldsymbol{x} = \boldsymbol{0}$.

***Fundamental Theorem of Linear Algebra*, Part 2**

$N(A)$ is the orthogonal complement of the row space $C(A^{\mathrm{T}})$ (in $\mathbf{R}^n$).

$N(A^{\mathrm{T}})$ is the orthogonal complement of the column space $C(A)$ (in $\mathbf{R}^m$).

$A^T A$

- Let's solve '$\boldsymbol{Ax} = \boldsymbol{b}$'
  - Sometimes m<n ← not easy to solve it
  - Sometimes m>n ← might be disturbed by noise due to too many equations or information

← $\boldsymbol{b}$ might not be in the subspace of C(A),

column space of A, then no solution

- Let's have '$A^T A$'
  - It's square
  - It's symmetric ← $(\boldsymbol{A^T A})^{\boldsymbol{T}} = \boldsymbol{A^T A}$
  - Is it invertible??

Key : let's solve $\boldsymbol{A^T A\hat{x}} = \boldsymbol{A}^{\mathbf{T}}\mathbf{b}$

* We hope this form would have solution even if $\boldsymbol{Ax} = \boldsymbol{b}$ doesn't, if $\boldsymbol{A^T A}$ is ***invertible***

BCML Bio Computing & Machine Learning Lab

$A^T A$

- Example

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

do we always expect to solve $\boldsymbol{Ax} = \boldsymbol{b}$**?**

No because **b** might not be in the column space of A, **C(A)**

➔ $\boldsymbol{A^T A} = \begin{bmatrix} \mathbf{3} & \mathbf{8} \\ \mathbf{8} & \mathbf{30} \end{bmatrix}$ is invertible.

But what if we have $A_2 = \begin{bmatrix} 1 & 3 \\ 1 & 3 \\ 1 & 3 \end{bmatrix}$, $\boldsymbol{A_2^T A_2} = \begin{bmatrix} \mathbf{3} & \mathbf{9} \\ \mathbf{9} & \mathbf{27} \end{bmatrix}$ is not invertible

$\boldsymbol{A_2}$'s rank is 1

$A^T A$

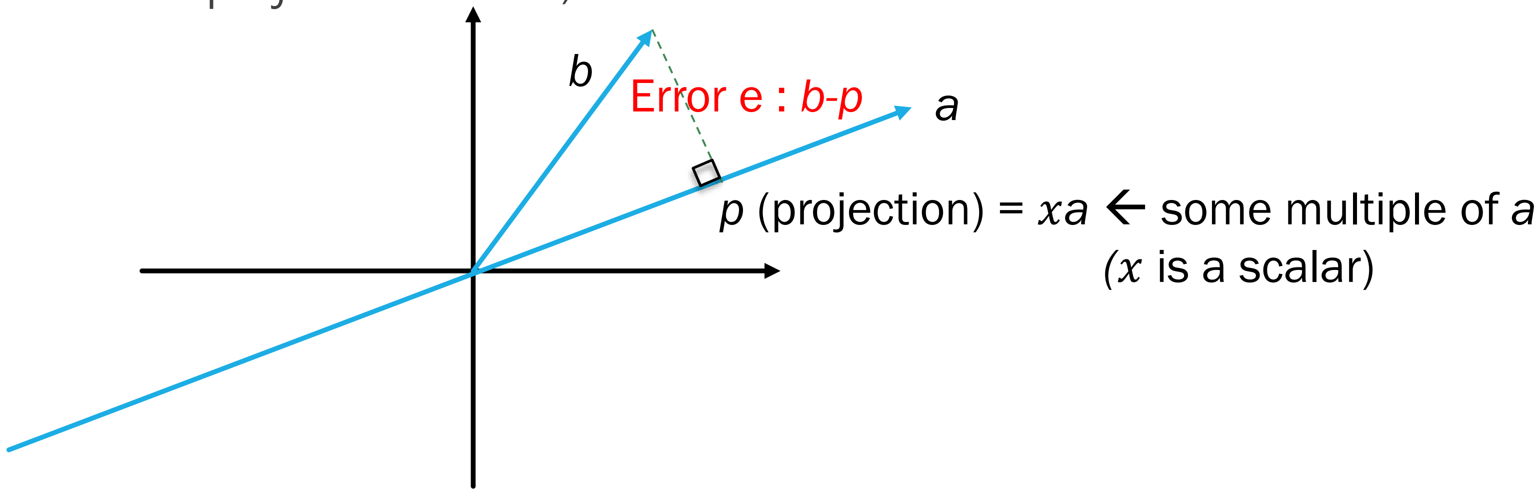
- $\boldsymbol{N(A^T A) = N(A)}$
- Rank of $\boldsymbol{A^T A}$ = rank of **A**

➔ $\boldsymbol{A^T A}$ is invertible exactly if **N(A)** only got the zero vector!

(this means A has independent columns)

PROJECTION ONTO A LINE

- One dimensional projection onto *a, which is one dimensional line*



$$a^T(b - xa) = 0 \rightarrow xa^T a = a^T b \rightarrow x = \frac{a^T b}{a^T a}$$

$$p = xa = a\frac{a^T b}{a^T a}$$

PROJECTION ONTO A LINE

- Projection matrix P

$$p = xa = a\frac{a^Tb}{a^Ta}$$

$$\text{Projection matrix : } P = \frac{aa^T}{a^Ta}$$

n X n matrix

scalar

- Projection 'p' vector is in the column space of P, C(P), based on the components of b vector

$$projection\ p = Pb$$

- Column space of P, C(P), is a line through 'a', which is one dimensional line

  ➔ rank(P) = 1 → column is the basis, one dimension

PROJECTION ONTO A LINE

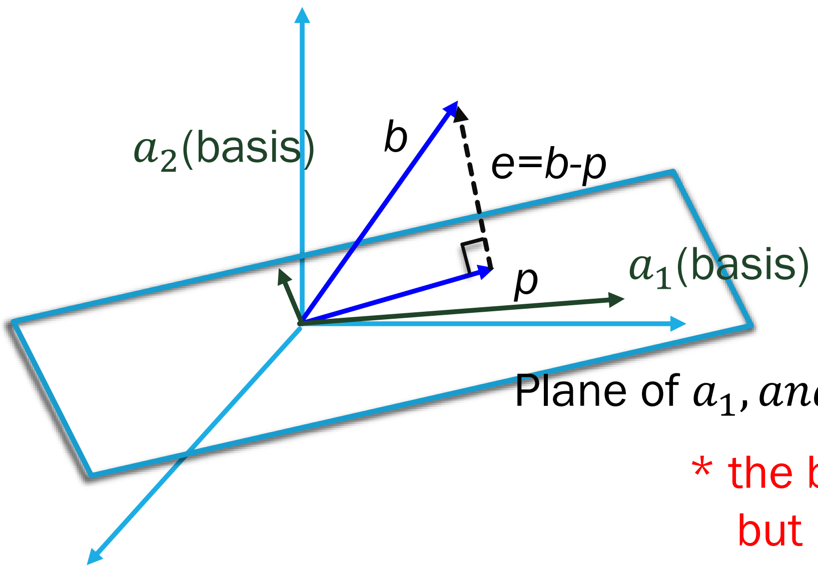
- *Two properties of projection matrix, P*
  - $P = \frac{aa^T}{a^Ta}$ is symmetric ← $P^T = P$
  - P^2 (projection twice) = P (projected onto the same place, p)

$$\frac{aa^Taa^T}{a^Taa^Ta} = \frac{a|a|^2a^T}{|a|^2|a|^2} = \frac{aa^T}{a^Ta}$$

PROJECTIONS

- Why project?
  - Because $Ax = b$ may have no solution, which means b is not in C(A)
  - Solve $A\hat{x} = p$, where p is a projection of b onto the column space of A, C(A)

PROJECTIONS ONTO A SUBSPACE



Now, we are considering 3 dimensional matrix of A
,while the previous example was one dimensional vector a

Plane of $a_1, and\ a_2$ from a column space of A, C(A)

* the bases don't have to be perpendicular each other,
but independent (no zero with their combination)

- Plane of $a_1 and\ a_2$ = column space of $A = [a_1\ a_2]$
- e is perpendicular to the plane
- $p = \widehat{x_1} a_1 + \widehat{x_2}\, a_2 = A\hat{x}$, find $\hat{x}$

key : $b - A\hat{x}$ is perpendicular to the plane →perpend. to $a_1\ and\ a_2$

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

PROJECTIONS ONTO A SUBSPACE

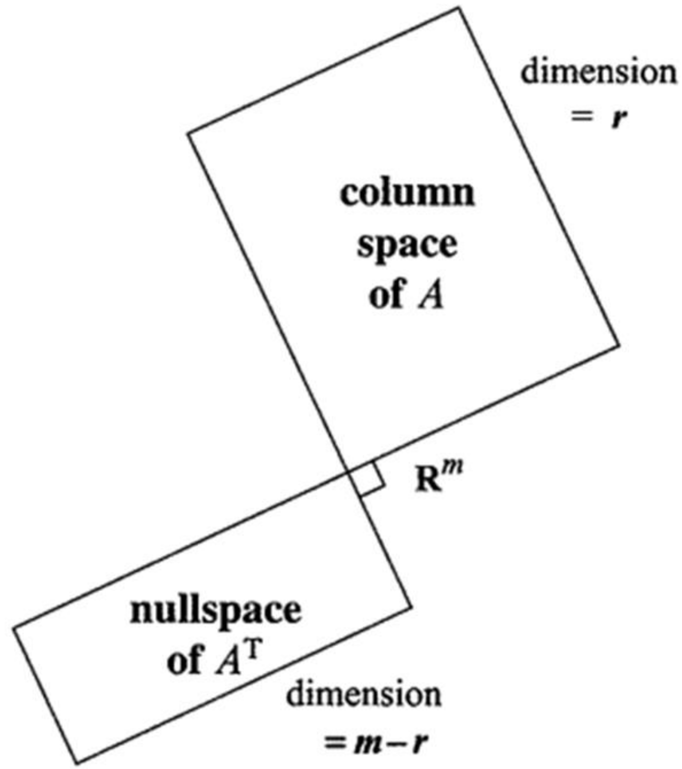
- $a_1^T(b - A\hat{x})=0$ and $a_2^T(b - A\hat{x})=0$

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

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

➔ $'e'$ is in $N(A^T)$ → $e \perp C(A)$

- $A^TA\hat{x} = A^Tb$



* we cannot just say $A^TA\hat{x} = A^Tb$ → $A\hat{x} = b$
since A^T is not always invertible.
If then, $A^T(b - A\hat{x})=0$ → $b - A\hat{x} = 0 \rightarrow A\hat{x} = b \rightarrow \mathrm{e} = 0$ means
b is on a column space of A already

PROJECTIONS ONTO A SUBSPACE

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

  → $\hat{x} = (A^T A)^{-1} A^T b$

  $p = \widehat{x_1} a_1 + \widehat{x_2}\, a_2 = A\hat{x}$ → $p = A(A^T A)^{-1} A^T b$

  projection matrix $P = A(A^T A)^{-1} A^T$

- $P^T = P$
- $P^2 = P$ ← $A(A^T A)^{-1} \cancel{A^T A (A^T A)^{-1}} A^T$

* When A is not a square matrix, we cannot always expect A^{-1}. Thus $A(A^T A)^{-1} A^T$ is not always 'I'.

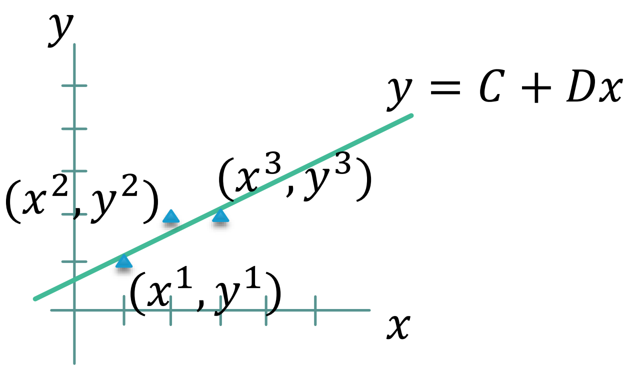
PROJECTIONS ONTO A SUBSPACE

- Example: Least Square
  - We hope

$$C + D = 1$$

$$C + 2D = 2$$

$$C + 3D = 2$$



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

We may not solve this equation, but

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

Since $\hat{x} = (A^T A)^{-1} A^T b$ from the previous slide, we can solve it

PROJECTIONS ONTO A SUBSPACE

- Projection Matrix : $P = A(A^TA)^{-1}A^T$
  - If *b* in column space, then *Pb=b*
  - If *b* ⊥ column space, then *Pb=0*
- Column space is perpendicular

  to the null space of A^T

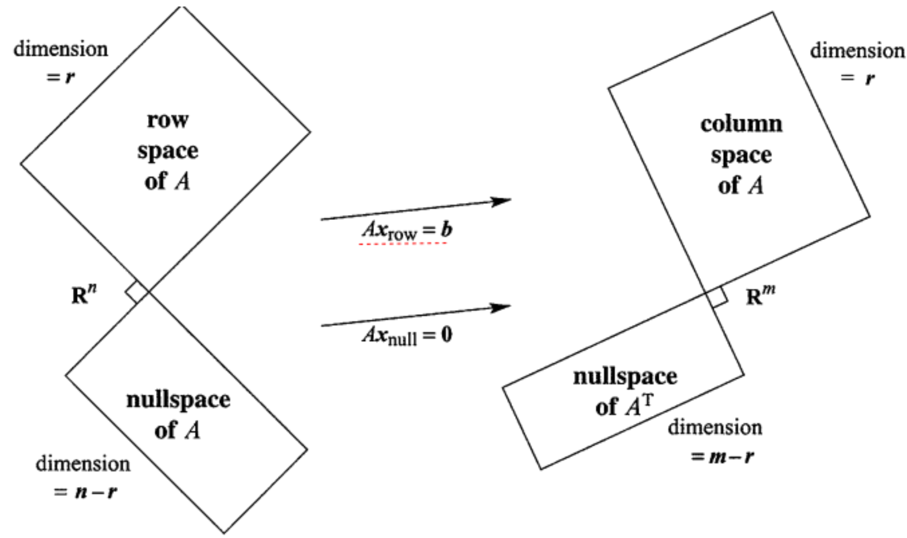
→ If *b* ⊥ column space, $A^Tb = 0$

$$A^T y = \begin{bmatrix} (\text{column } 1)^T \\ \cdots \\ (\text{column } n)^T \end{bmatrix} \begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} 0 \\ \cdot \\ 0 \end{bmatrix}.$$

→ $p = Pb = A(A^TA)^{-1}\underline{A^Tb = 0}$

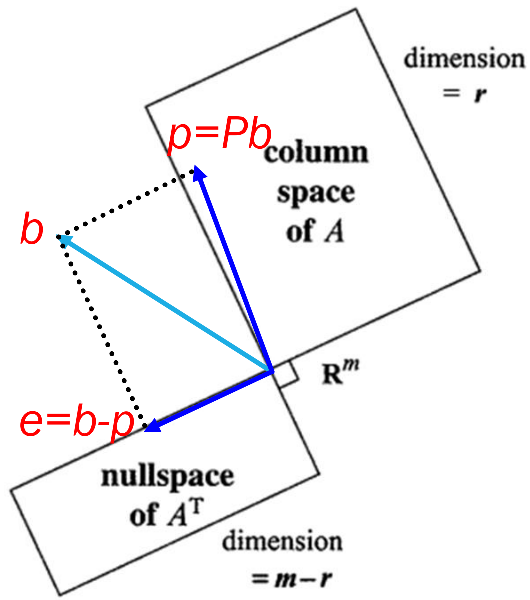
→ If *b* in column space, <u>*b=Ax*</u>

→ $p = Pb = A\cancel{(A^TA)^{-1}A^T}Ax = \underline{\mathrm{Ax} = \mathrm{b}}$



PROJECTIONS ONTO A SUBSPACE

- '*p*' is a projection of *b* onto column space
  p=Pb
- Error vector, '*e=b-p*', should be orthogonal to column space due to the projection property
  → '*e*' is on the nullspace of A^T!
- $e = b - p = b - Pb = (I - P)b$
- If P is a projection, then $I - P$ is also projection
  Projections onto the nullspace of A^T
- If P is symmetric, $I - P$ is symmetric
- If P^2 is equal to P, then $(I - P)^2$ equals $I - P$



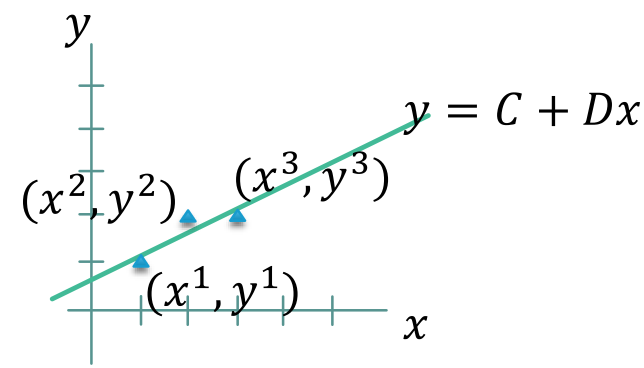
LEAST SQUARE APPROXIMATIONS

- Example: Least Square
  - We hope

$$C + D = 1$$

$$C + 2D = 2$$

$$C + 3D = 2$$



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

We may not solve this equation, but

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

Since $\hat{x} = (A^T A)^{-1} A^T b$ from the previous slide, we can solve it

LEAST SQUARE APPROXIMATIONS

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

$$\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$$

$$\rightarrow D = \frac{1}{2}, C = \frac{2}{3}$$

- Use partial derivative

$$Minimize\ E = |Ax - b|^2 = |e|^2 = e_1^2 + e_2^2 + e_3^2$$
$$= (C + D - 1)^2 + (C + 2D - 2)^2 + (C + 3D - 2)^2$$

$$\frac{dE}{dC} = 0\ and\ \frac{dE}{dD} = 0$$

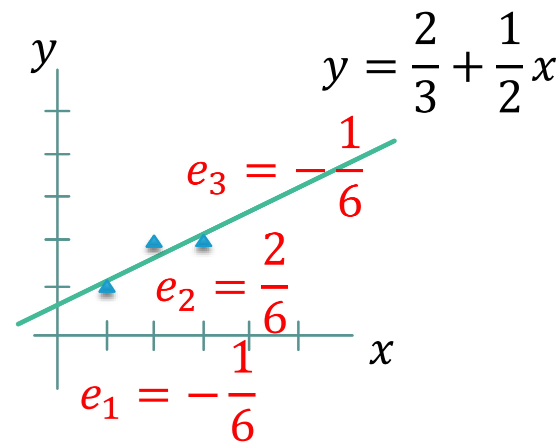
- Best line : $\frac{2}{3} + \frac{1}{2} t$

BCML Bio Computing & Machine Learning Lab

LEAST SQUARE APPROXIMATIONS

- $b = (1,2,2)$

- $p = \overset{\color{red}A}{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}} \overset{\color{red}\hat{x}}{\begin{bmatrix} \frac{2}{3} \\ \frac{1}{2} \end{bmatrix}} = \begin{bmatrix} \frac{7}{6} \\ \frac{5}{3} \\ \frac{13}{6} \end{bmatrix}$



- $e = b - p = \left(1 - \frac{7}{6}; 2 - \frac{5}{3}; 2 - \frac{13}{6}\right) = (-\frac{1}{6}; \frac{1}{3}; -\frac{1}{6})$
- $p \perp e = p^T \cdot e = 0$

- $e \perp C(A) = \left[-\frac{1}{6}, \frac{1}{3}, -\frac{1}{6}\right] \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = [0, 0]$

LEAST SQUARE APPROXIMATIONS

- If A has independent columns, then $A^T A$ is invertible

prove) A has independent columns and let's find the nullspace of $A^T A$

$$A^T Ax = 0$$

Let's multiply x^T on both sides of $A^T Ax = 0$

$$x^T A^T Ax = 0 \rightarrow (Ax)^T (Ax) = |Ax|^2 = 0$$

the length of Ax, $|Ax|$, is zero, and so $Ax = 0$

if A has independent columns, then its nullspace should be zero $(x = 0)$

because there is no free column

Thus, the nullspace of $A^T A$ is also zero, $x = 0$

Now we can say that

if A (square or non-square, whatever) has independent columns,

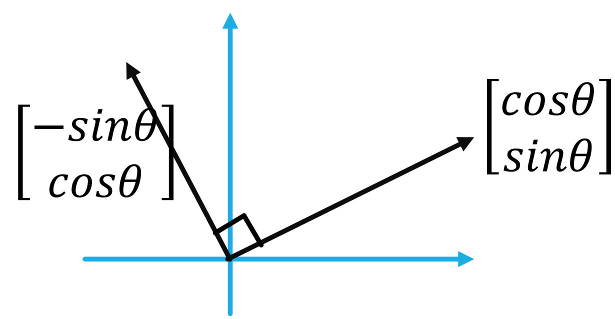
then $A^T A$ is invertible since its nullspace is zero

ORTHOGONAL BASES

- Columns are ***definitely independent*** if they are perpendicular unit vectors

examples) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

➔ Orthonomal vectors are the best columns



ORTHOGONAL BASES

- Ortho-normal Vectors

$$q_i^T q_j = \begin{cases} 0, if\ i \neq j \\ 1, if\ i = j \end{cases}$$

- $Q = [q_1, \dots, q_n]$, $\quad Q^T Q = \begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix} [q_1 \ \dots q_n] = I$

- If an orthogonal matrix Q *is square*, then $Q^T Q = I \rightarrow Q^T = Q^{-1}$ since Q is invertible, which has independent columns

Example) permutation matrix $Q = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = I$

ORTHOGONAL BASES

Example) $\begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}$ is orthogonal(normal) matrix

Example) $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ is orthogonal(normal) matrix

Example) $\frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$ is orthogonal(normal) matrix

ORTHOGONAL BASES

- Q has orthonormal columns, and then its projection matrix

$$P = Q\underbrace{(Q^TQ)}_{\text{Identity}}{}^{-1}Q^T = QQ^T = \begin{cases} I \ \textit{if } Q \text{ is square } (Q^T = Q^{-1}) \\ \text{not } I \text{ if } Q \text{ is not square} \end{cases}$$

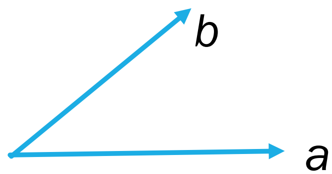
- P is a symmetric matrix ← QQ^T is symmetric
- $P^2 = P$ → $(Q\underbrace{Q^T)(Q}_{\text{Identity}}Q^T) = QQ^T$

- $A^TA\hat{x} = A^Tb \rightarrow Q^TQ\hat{x} = Q^Tb$

$$\Rightarrow \hat{x} = Q^Tb \rightarrow \boxed{\hat{x_i} = q_i^Tb}$$

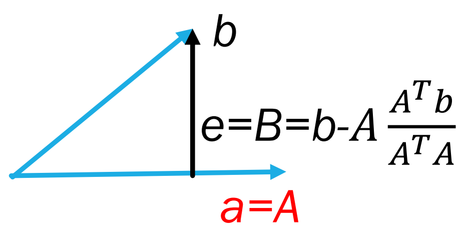
GRAM-SCHMIDT

- Make the columns orthonormal
- Independent vectors *a* and *b* → orthogonal A,B



→ orthonormal $q_1 = \frac{A}{|A|}$, $q_2 = \frac{B}{|B|}$

- Make *a* and *b* orthogonal, *A* and *B*

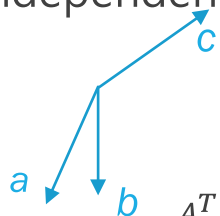


$A \perp B$

- $A^T B = A^T \left(b - A \frac{A^T b}{A^T A} \right) = 0$

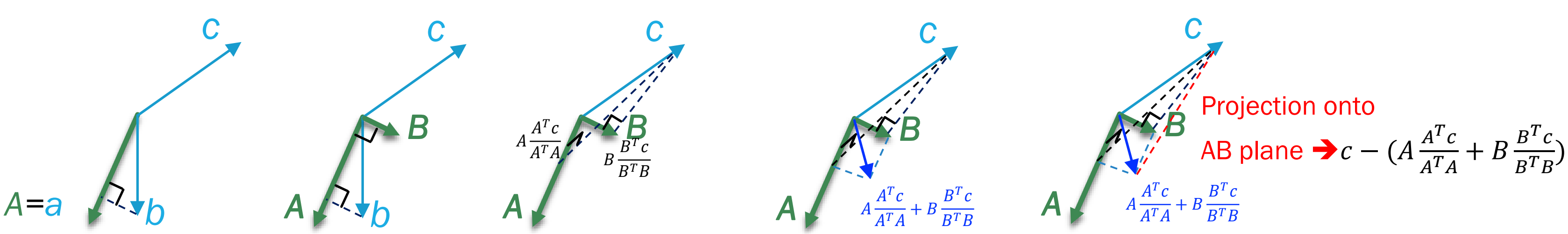
GRAM-SCHMIDT

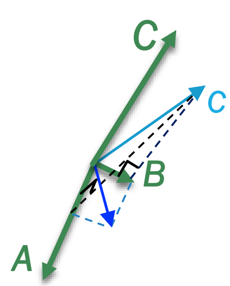
- Independent vectors *a, b* and *c* → orthogonal A,B and C



→ orthonormal $q_1 = \frac{A}{|A|},\ q_2 = \frac{B}{|B|}$ and $q_3 = \frac{C}{|C|}$

- $C = c - A\frac{A^Tc}{A^TA} - B\frac{B^Tc}{B^TB} \ \Rightarrow\ C \perp A,\ \ C \perp\ B$





GRAM-SCHMIDT

- $a = \begin{bmatrix}1\\1\\1\end{bmatrix}, b = \begin{bmatrix}1\\0\\2\end{bmatrix} \Rightarrow B = \begin{bmatrix}1\\0\\2\end{bmatrix} - \begin{bmatrix}1\\1\\1\end{bmatrix}\frac{A^T b}{A^T A} = \begin{bmatrix}1\\0\\2\end{bmatrix} - \begin{bmatrix}1\\1\\1\end{bmatrix}\frac{3}{3} = \begin{bmatrix}0\\-1\\1\end{bmatrix}$

A ⊥ B

- $Q = [q_1 \;\; q_2] = \begin{bmatrix}\frac{1}{\sqrt{3}} & 0\\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}}\end{bmatrix}$ ← Orthonormal (Schmidt algorithm)

GRAM-SCHMIDT

- The Factorization $A = QR$

- The vectors $\boldsymbol{a}$ and $\boldsymbol{A}$ and $\boldsymbol{q}_1$ are all along a single line.
- The vectors $\boldsymbol{a}, \boldsymbol{b}$ and $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{q}_1, \boldsymbol{q}_2$ are all in the same plane.
- The vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ and $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ and $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ are in one subspace (dimension 3).

At every step $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ are combinations of $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_k$. Later $\boldsymbol{q}$'s are not involved. The connecting matrix R is ***triangular***, and we have $A = QR$:

$$\begin{bmatrix} \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{c} \\ & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{c} \\ & & \boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{c} \end{bmatrix} \quad \text{or} \quad A = QR. \tag{9}$$

$A = QR$ is Gram-Schmidt in a nutshell. Multiply by Q^{T} to see why $R = Q^{\mathrm{T}}A$.

BCML Bio Computing & Machine Learning Lab