Math 2270-1

Notes of 11/19/19

Announcements

6. Orthogonality and Least Squares

- The **inner product**, previously called the **dot product**, of two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, is defined to be

$$\mathbf{u}\bullet\mathbf{v} = \mathbf{u}^T\mathbf{v} = [\, u_1 \quad u_2 \quad \ldots \quad u_n \,] \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \sum_{i=1}^{n} u_i v_i. \tag{1}$$

- **Theorem 1, p. 333**. Let $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ be vectors in $\mathbb{R}^n$, and c be a scalar. Then

  a. $\mathbf{u} \bullet \mathbf{v} = \mathbf{v} \bullet \mathbf{u}$

  b. $(\mathbf{u} + \mathbf{v}) \bullet \mathbf{w} = \mathbf{u} \bullet \mathbf{w} + \mathbf{v} \bullet \mathbf{w}$

  c. $(c\mathbf{u}) \bullet \mathbf{v} = c(\mathbf{u} \bullet \mathbf{v})$

  d. $\mathbf{u} \bullet \mathbf{u} \geq 0, \quad \text{and} \quad \mathbf{u} \bullet \mathbf{u} = 0 \Longrightarrow \mathbf{u} = \mathbf{0}.$

- The **length** or **norm**[−1−] of a vector $\mathbf{v}$ is defined by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \bullet \mathbf{v}}. \tag{2}$$

[−1−] also called **Standard Norm**, **Euclidean Norm**, or **2-norm**.

- Definition: Two vectors $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** (or **perpendicular**) if

$$\mathbf{u} \bullet \mathbf{v} = 0. \tag{3}$$

the zero vector is orthogonal to all vectors in $\mathbb{R}^n$.

- Suppose W is a subspace of $\mathbb{R}^n$. Then the set

$$W^\perp = \{\mathbf{z} : \mathbf{z} \text{ is orthogonal to all vectors in } W\} \tag{4}$$

is a linear space, called the **orthogonal complement** of W.

- $W^\perp$ is read as "W-perpendicular" or, more commonly, just "W-perp".

- Example: line and plane in $\mathbb{R}^3$.

- **Theorem 3**, p. 337: Let A be an $m \times n$ matrix. The orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of the column space of A is the null space of A^T

$$(\mathrm{Row}A)^\perp = \mathrm{Nul}A \quad \text{and} \quad (\mathrm{Col}A)^\perp = \mathrm{Nul}A^T. \tag{5}$$

- A set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ from $\mathbb{R}^n$ is an **orthogonal set** if each pair of distinct vectors from that set is orthogonal, i.e.,

$$i \neq j \quad \Longrightarrow \quad \mathbf{u}_i \bullet \mathbf{u}_j = 0. \tag{6}$$

- **Theorem 4**, p. 340, textbook. If

$$S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\} \tag{7}$$

is an orthogonal set of **nonzero** vectors in $\mathbb{R}^n$, then S is linearly independent. (Hence S is a basis of span(S).)

- Naturally, an **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

- Orthogonal Bases are nice! You can compute coefficients without solving a linear system.

- Suppose

$$\mathrm{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\} \tag{8}$$

is a basis of a subspace W of $\mathbb{R}^n$,

$$B = [\,\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\,], \tag{9}$$

and $\mathbf{y}$ is a vector in W. Then, in general, computing the coordinate vector

$$[\mathbf{y}]_{\mathrm{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix} \tag{10}$$

of $\mathbf{y}$ requires the solution of the linear system

$$B[\mathbf{y}]_{\mathrm{B}} = \mathbf{y}. \tag{11}$$

- However, if B is an orthogonal basis we can compute the components of $[\mathbf{y}]_{\mathrm{B}}$ directly:

$$c_j = \frac{\mathbf{y} \bullet \mathbf{u}_j}{\mathbf{u}_j \bullet \mathbf{u}_j}. \tag{12}$$

- **Theorem 6**, p. 345. An $m \times n$ matrix U has orthonormal columns if and only if

$$U^T U = I \tag{13}$$

(where I is the $n \times n$ identity matrix.).

- **Theorem 7**, p. 345. Let U be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}$ and y be vectors in $\mathbb{R}^n$. Then:

  a. $\|U\mathbf{x}\| = \|\mathbf{x}\|$

  b. $(U\mathbf{x}) \bullet (U\mathbf{y}) = \mathbf{x} \bullet \mathbf{y}$

  c. $(U\mathbf{x}) \bullet (U\mathbf{y}) = 0 \quad$ if and only if $\quad \mathbf{x} \bullet y = 0$

- The Pythagorean Theorem states that

$$\|\mathbf{u} \pm \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \quad \Longleftrightarrow \quad \mathbf{u} \bullet \mathbf{v} = 0. \tag{14}$$

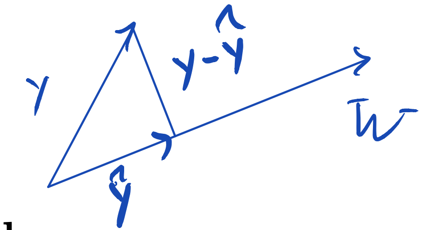
- The **orthogonal projection** of a vector $\mathbf{v}$ onto a vector $\mathbf{u}$ is given by

$$\mathrm{proj}_{\mathbf{u}}\mathbf{v} = \frac{\mathbf{u} \bullet \mathbf{v}}{\mathbf{u}\mathbf{u}}\mathbf{u}. \tag{15}$$

- **Theorem 8.** Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \tag{16}$$

where $\hat{\mathbf{y}}$ is in W and $\mathbf{z}$ is in $W^\perp$.



- This is the **orthogonal Decomposition theorem.** The vector $\hat{\mathbf{y}}$ in (16) is called the **orthogonal projection of y onto** W.

$W = \text{span}\{w\}$

$\hat{y} = \dfrac{y \cdot w}{w \cdot w} w$

- The textbook uses the notation

$$\hat{\mathbf{y}} = \text{proj}_W \mathbf{y}. \tag{17}$$

- **Best Approximation Theorem** (Theorem 9, p. 352) Let W be a subspace of $\mathbb{R}^n$, let $\mathbf{y}$ be any vector in $\mathbb{R}^n$, and let $\hat{\mathbf{y}}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then $\hat{\mathbf{y}}$ is the closest point W to $\mathbf{y}$, in the sense that

$$\|\mathbf{y} - \mathbf{v}\| \overset{>}{\cancel{<}} \|\mathbf{y} - \hat{\mathbf{y}}\| \tag{18}$$

for all $\mathbf{v}$ in W distinct from $\hat{\mathbf{y}}$.

- **Theorem 10**, p. 353. If $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$\hat{\mathbf{y}} = \text{proj}_W \mathbf{y} = \sum_{i=1}^{p} (\mathbf{y} \bullet \mathbf{u}_i)\mathbf{u}_i. \tag{19}$$

If $U = [\,\mathbf{u}_1 \quad \mathbf{u})2 \quad \ldots \mathbf{u}_p\,]$, then

$$\text{proj}_W \mathbf{y} = UU^T \mathbf{y} \tag{20}$$

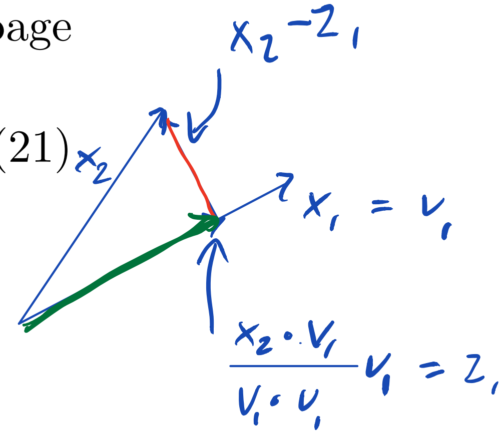
for all $\mathbf{y}$ in $\mathbb{R}^n$.

- We considered three versions of the Gram-Schmidt Process.

- Version 1: is described by **Theorem 11**, page 357, textbook: Given a basis

$$\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\} \tag{21}$$

for a non-zero subspace W of $\mathbb{R}^n$, define



$$
\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \bullet \mathbf{v}_1}{\mathbf{v}_1 \bullet \mathbf{v}_1}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \bullet \mathbf{v}_1}{\mathbf{v}_1 \bullet \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \bullet \mathbf{v}_2}{\mathbf{v}_2 \bullet \mathbf{v}_2}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p \bullet \mathbf{v}_1}{\mathbf{v}_1 \bullet \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \bullet \mathbf{v}_2}{\mathbf{v}_2 \bullet \mathbf{v}_2}\mathbf{v}_2 - \ldots - \frac{\mathbf{x}_p \bullet \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \bullet \mathbf{v}_{p-1}}\mathbf{v}_{p-1}
\end{aligned} \tag{22}
$$

Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is an orthogonal basis for W. In addition,

$$\text{span}\,\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \text{span}\,\{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \quad \text{for} \quad k = 1, 2, \ldots, p. \tag{23}$$

- Version 2 is just a more compact notation for the process. For $k = 1, \ldots, p$ define

$$\mathbf{v}_k = \mathbf{x}_k - \sum_{i=1}^{k-1} \frac{\mathbf{x}_k \bullet \mathbf{v}_i}{\mathbf{v}_i \bullet \mathbf{v}_i}\mathbf{v}_i. \tag{24}$$

- Version 3 combines normalization with orthogonalization: For $k = 1, \ldots, p$ define

$$\begin{cases} \mathbf{w}_k &= \mathbf{x}_k - \displaystyle\sum_{i=1}^{k-1} (\mathbf{x}_k \bullet \mathbf{v}_i)\mathbf{v}_i \\ \mathbf{v}_k &= \dfrac{\mathbf{w}_k}{\|\mathbf{w}_k\|} \end{cases} \tag{25}$$

$\text{span}\{x_1, \ldots, x_k\} = \text{span}\{v_1, \ldots, v_k\}$
$k = 1, 2, \ldots, p$

- Definition: A square matrix Q is **orthogonal** if its columns form an **orthonormal** set.

- This means that

$$Q^T Q = I, \tag{26}$$

i.e., Q is invertible, and

$$Q^{-1} = Q^T. \tag{27}$$

(see textbook, page 346.)

- **Theorem 12**, page 359, textbook. If A is an~~d~~ $m \times n$ matrix with linearly independent columns, then A can be factored a

$$A = QR \tag{28}$$

$$\begin{bmatrix} \times & \ddots & \times \\ 0 & & \times \\ & 0 & \ddots \end{bmatrix} = R$$

where Q is an $m \times n$ matrix whose columns form an orthonormal basis for $\text{Col}(A)$ and R is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

- Suppose we have an overdetermined linear system

$$A\mathbf{x} = \mathbf{b} \tag{29}$$

- Here A is $m \times n$, $\mathbf{x}$ is in $\mathbb{R}^n$, $\mathbf{b}$ is in $\mathbb{R}^m$, and $m \geq n$ (and typically, $m > n$).

- Usually, the system (29) will not have a solution. In that case, the next best thing is to solve the alternative problem

$$\|A\mathbf{x} - \mathbf{b}\| = \min \tag{30}$$

- In other words (the words of our textbook), we want to find a vector $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\| \tag{31}$$

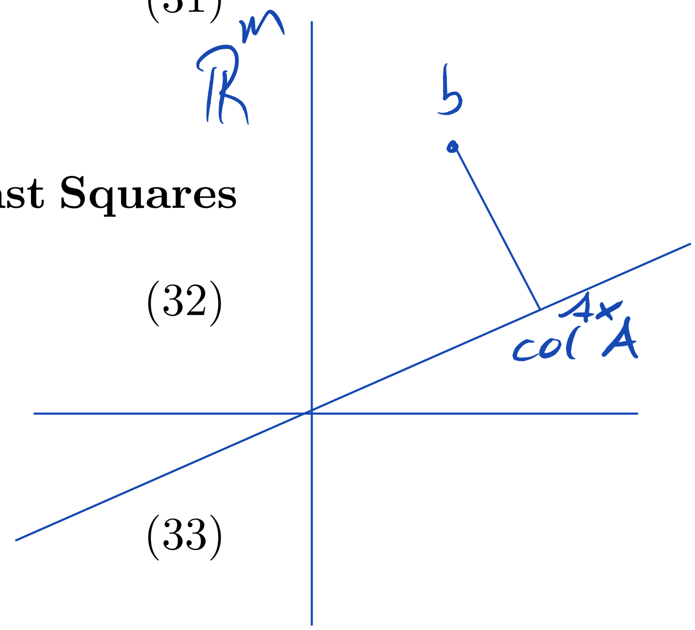
  for all $\mathbf{x}$ in $\mathbb{R}^n$.

- The textbook calls such an $\hat{\mathbf{x}}$ a **Least Squares Solution** of

$$A\mathbf{x} = \mathbf{b}. \tag{32}$$

- I would call it a solution of

$$\|A\mathbf{x} - \mathbf{b}\| = \min. \tag{33}$$



- First: **Theorem 13** (p. 363) The set of least square solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the **normal equations**

$$A^T A\mathbf{x} = A^T\mathbf{b}. \tag{34}$$

- **Theorem 14** (p. 365) Let A be an $m \times n$ matrix. The following statements are logically equivalent. (This means they are either all true or all false):

a. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.

b. The columns of A are linearly independent.

c. The matrix $A^T A$ is invertible.

- Suppose we write

$$\boxed{A = QR} \tag{35}$$

where

$$Q = \begin{array}{cc} & \begin{array}{cc} n & m-n \end{array} \\ m & \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \end{array} \tag{36}$$

is *orthogonal* and

$$R = \begin{array}{cc} & n \\ \begin{array}{c} n \\ m-n \end{array} & \begin{pmatrix} R_1 \\ 0 \end{pmatrix} \end{array} \tag{37}$$

with R_1 being upper triangular.

- Earlier we discussed how to obtain

$$A = Q_1 R_1, \tag{38}$$

for example by the Gram-Schmidt Process.

- To get Q from Q_1 we simply add vectors to the orthonormal basis of the column space of A to get an orthonormal bass of $\mathbb{R}^m$.

- We won't actually need Q_2, but it's useful to describe the idea.

- A significant property of an orthogonal matrix is that multiplying with it does not alter the norm of a vector:

$$\|Qx\|^2 = (Qx)^T(Qx) = x^T Q^T Q x = x^T x = \|x\|^2. \tag{39}$$

- Using

$$A = QR \quad \text{and} \quad Q^T A = R \tag{40}$$

  we obtain

$$\begin{aligned} \|Ax - b\|^2 &= \|Q^T(Ax - b)\|^2 \\ &= \|Q^T Ax - Q^T b\|^2 \\ &= \left\| \begin{pmatrix} R_1 x \\ 0 \end{pmatrix} - \begin{pmatrix} Q_1^T b \\ Q_2^T b \end{pmatrix} \right\|^2 \\ &= \left\| R_1 x - Q_1^T b \right\|^2 + \left\| Q_2^T b \right\|^2 . \end{aligned} \tag{41}$$

- Of the two terms on the right we have no control over the second, and we can render the first one zero by solving (the square triangular $n \times n$ linear system)

$$R_1 x = Q_1^T b. \tag{42}$$

- **Definition** (p. 378, textbook): An **inner product** on a vector space V is a function that, to each pair of vectors $\mathbf{u}$ and $\mathbf{v}$ in V, associates a real number $< \mathbf{u}, \mathbf{v} >$ and satisfies the following axioms, for all vectors $\mathbf{u}$ and $\mathbf{v}$ in V and all scalars c:

1. $< \mathbf{u}, \mathbf{v} > = < \mathbf{v}, \mathbf{u} >$.

2. $< \mathbf{u} + \mathbf{v}, \mathbf{w} > = < \mathbf{u}, \mathbf{w} > + < \mathbf{v}, \mathbf{w} >$

3. $< c\mathbf{u}, \mathbf{v} > = c < \mathbf{u}, \mathbf{v} >$

4. $< \mathbf{u}, \mathbf{u} > \geq 0 \quad$ and $\quad < \mathbf{u}, \mathbf{u} > = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

- A Vector space with an inner product is called an **inner product space**.
- The Cauchy-Schwarz Inequality says

$$| < \mathbf{u}, \mathbf{v} > | \leq \|\mathbf{u}\|\|\mathbf{v}\| \tag{43}$$

- The triangle inequality says

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|. \tag{44}$$

- One major application of inner product spaces is **weighted least squares**.
- The underlying space is $\mathbb{R}^n$ and the inner product is

$$< \mathbf{x}, \mathbf{y} >= \sum_{i=1}^{n} w_i x_i y_i \tag{45}$$

where the w_i are given positive weights.

- The normal equations for the weighted Least Squares Solution of

$$A\mathbf{x} = \mathbf{b} \tag{46}$$

are

$$A^T W A\mathbf{x} = A^T W \mathbf{b}. \tag{47}$$

- Another major example is **Fourier Series**. The underlying linear space is the set of 2π

$\langle f, g \rangle = \int_a^b f(x)g(x)dx$

periodic functions that are square integrable over an interval of length 2π.

- The underlying inner product is

$$< f, g >= \int_{-\pi}^{\pi} f(t)g(t)\mathrm{d}t. \tag{48}$$

- The Fourier series of a 2π-periodic function f is

$$f(\mathbf{t}) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(n\mathbf{t}) + b_n \sin(nt) \tag{49}$$

where the **Fourier coefficients** are given by

$$\begin{aligned} a_n &= \frac{< f, \cos(nt) >}{\pi} = \frac{\int_{-\pi}^{\pi} f(t)\cos(nt)\mathrm{d}t}{\pi} \\ b_n &= \frac{< f, \sin(nt) >}{\pi} = \frac{\int_{-\pi}^{\pi} f(t)\sin(nt)\mathrm{d}t}{\pi} \end{aligned} \tag{50}$$

$$\int_{-\pi}^{\pi} \cos^2 nt\,dt = \frac{1}{2}\int_{-\pi}^{\pi} \sin^2 nt + \cos^2 nt\,dt$$

$$= \frac{1}{2}\int_{-\pi}^{\pi} 1\,dt = \frac{1}{2}2\pi = \pi$$

$$v_1 = \begin{bmatrix} 1 \\ 2 \\ 7 \end{bmatrix} \qquad v_2 \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}$$

$$u_1 = \frac{\begin{bmatrix} 1 \\ 2 \\ 7 \end{bmatrix}}{\sqrt{54}}$$

$$w_2 = \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix} - \frac{9}{\sqrt{54}} \cdot \frac{1}{\sqrt{54}} \begin{bmatrix} 1 \\ 2 \\ 7 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix} - \frac{1}{6} \begin{bmatrix} 1 \\ 2 \\ 7 \end{bmatrix} = \begin{bmatrix} 5/6 \\ -10/3 \\ -5/6 \end{bmatrix}$$

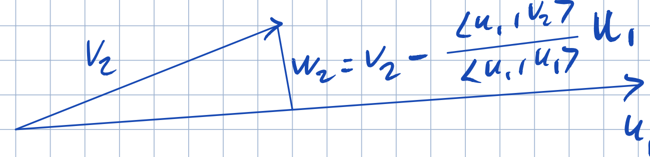
$$\{v_1, v_2, v_3, \ldots\}$$

$$u_2 = \frac{\begin{bmatrix} 5/6 \\ -10/3 \\ -5/6 \end{bmatrix}}{\sqrt{\frac{25}{36} + \frac{100}{9} + \frac{25}{36}}}$$

$$\{u_1, u_2, u_3, \ldots\}$$

$$\langle u_i, u_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

$$u_1 = \frac{v_1}{\|v_1\|}$$



$$w_2 = v_2 - \frac{\langle u_1, v_2 \rangle}{\langle u_1, u_1 \rangle} u_1$$

$$= v_2 - \langle u_1, v_2 \rangle u_1$$

$$v_2 = \frac{w_2}{\|w_2\|}$$

$$\text{span}\{v_1, \ldots, v_k\} = \text{span}\{u_1, \ldots, u_k\}$$

$$k = 1, 2, 3, \ldots$$

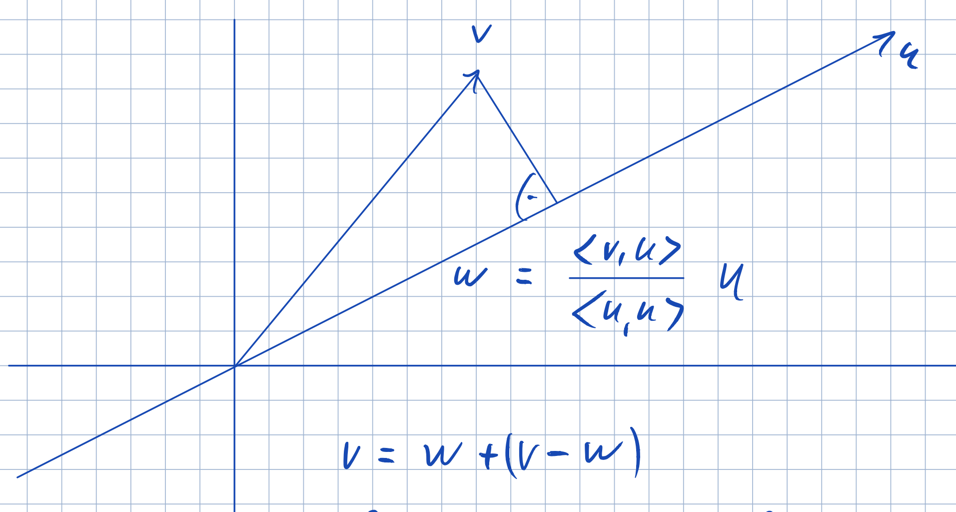
$$\begin{bmatrix} 1 & 1 \\ 2 & -3 \\ 7 & 2 \end{bmatrix} = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \overset{R}{\begin{bmatrix} \times & \times \\ 0 & \times \end{bmatrix}}$$

$$A = QR \qquad Q^TQ = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} u_1^T \\ u_2^T \end{bmatrix} \begin{bmatrix} u_1 & u_2 \end{bmatrix} = \begin{bmatrix} u_1^T u_1 & u_1^T u_2 \\ u_2^T u_1 & u_2^T u_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$Q^TA = Q^TQR = R$$

CS $\quad |\langle u, v\rangle| \le \|u\| \, \|v\|$



$$w = \frac{\langle v, u\rangle}{\langle u, u\rangle} u$$

$$v = w + (v - w)$$

$$\|v\|^2 = \|w\|^2 + \|v - w\|^2$$

$$\|v\|^2 \geq \|w\|^2$$

$$\langle v, v\rangle \geq \left\langle \frac{\langle v, u\rangle}{\langle u, u\rangle} u, \frac{\langle v, u\rangle}{\langle u, u\rangle} u \right\rangle$$

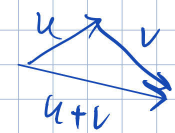
$$= \frac{\langle v, u\rangle^2}{\langle u, u\rangle^2} \langle u, u\rangle = \frac{\langle v, u\rangle^2}{\langle u, u\rangle}$$

$$\langle v, v\rangle \langle u, u\rangle \geq \langle v, u\rangle^2$$

$$\|v\|^2 \|u\|^2 \geq |\langle v, u\rangle|^2 \quad \text{Cauchy-Schwartz}$$

$$\|v\| \|u\| \geq |\langle v, u\rangle|$$

$$\|u+v\| \leq \|u\| + \|v\|$$



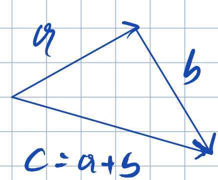
$$\|u+v\|^2 \leq (\|u\| + \|v\|)^2$$

$$\langle u+v, u+v\rangle = \langle u, u\rangle + 2\langle u, v\rangle + \langle v, v\rangle$$

$$\leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2$$

$$= (\|u\| + \|v\|)^2$$

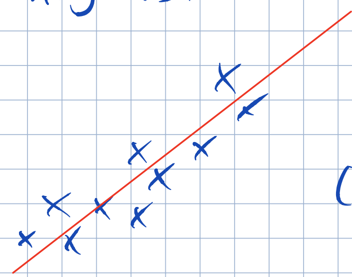
Pythagorean Thm



$$\|a+b\|^2 = \|a\|^2 + 2\langle a, b\rangle + \|v\|^2$$

Linear Regression

Regression = Least Squares



(x_i, y_i)

$$y = \alpha x + \beta$$

suppose they all lie on the line

$$y_i = \alpha x_i + \beta \cdot 1 \qquad i = 1 \dots$$

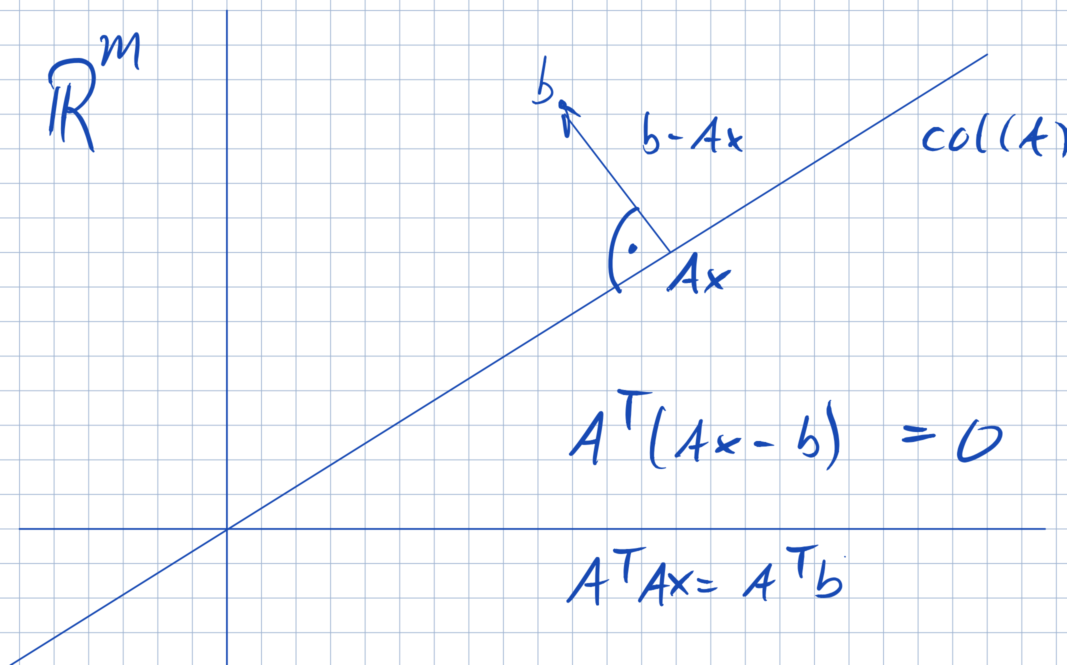
$$\begin{bmatrix} x_1 & 1 \\ | & | \\ x_m & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} y_1 \\ | \\ y_m \end{bmatrix}$$

$$Ax = b$$

$m \times n$

$$A^T A = \begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & \underbrace{\sum 1}_{=m} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \sum x_i y_i \\ \sum y_i \end{bmatrix}$$

$$\|Ax - b\| = \min$$



$$A^T(Ax - b) = 0$$

$$A^T A x = A^T b$$