Linear Algebra
Week 10

Felix Breuer
G-04

Last time we saw:

$$x^* = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \|Ax-b\|^2 \iff \underbrace{A^TAx^* = A^Tb}_{\text{normal equations}}$$

Single choice (only one option is correct):

- ☐ $b - Ax^*$ is orthogonal to the row space of A
- ☒ $b - Ax^*$ is orthogonal to the column space of A
- ☐ x^* is in the null space of A
- ☐ the solution x^* does not always exist

Let $A \in \mathbb{R}^{m \times n}$, $\operatorname{rank}(A) = n$:

$\Rightarrow A^TA \in \mathbb{R}^{n \times n}$, $\operatorname{rank}(A^TA) = n$

(see exercise 7.2, from $C(A^T) = C(A^TA)$ we can also follow that $A^TAx = A^Tb$ always has a solution)

Hence A^TA is invertible. This allows us to write:

$$\begin{aligned} A^TAx &= A^Tb \\ \iff \quad x &= (A^TA)^{-1}A^Tb \\ \iff \quad Ax &= A(A^TA)^{-1}A^Tb \end{aligned}$$

Ax is the orthogonal projection of b onto $C(A)$.

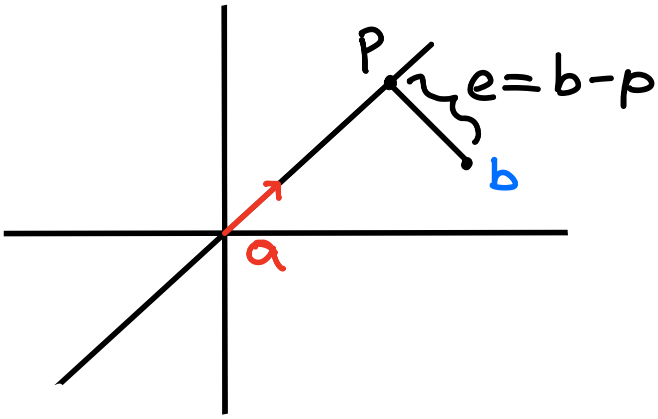
The projection matrix to project onto $C(A)$ is given by

$$\boxed{P = A(A^TA)^{-1}A^T}$$

Applying this formula for $A = a \in \mathbb{R}^n$, $a \neq 0$ gives us:

$$Pb = \operatorname{proj}_{\operatorname{span}(a)}(b) = a(a^Ta)^{-1}a^Tb = \frac{a^Tb}{a^Ta}a = \boxed{\frac{aa^T}{a^Ta}b}$$

Alternatively, we could derive this similarly to how we derived the normal equations:



$$b-p \perp a$$
$$\Rightarrow a^T(b-p) = 0$$
$$\Leftrightarrow a^T b = a^T p$$
$$\Leftrightarrow a^T b = a^T x a$$
$$\Leftrightarrow x = \frac{a^T b}{a^T a}$$
$$\Leftrightarrow ax = a \frac{a^T b}{a^T a}$$

Recommendation: Compare this to the geometric definition of the dot product from week two.

Orthonormality

We call a set of vectors $\{q_1, \dots, q_n\}$ orthonormal if each vector in the set has length one and is orthogonal to all others in the set. We can express this as:

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

1) • Can you give an example of such a set?
2) • Is any such set linearly independent?

1) An example of such a set is the canonical basis of $\mathbb{R}^3$,
$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\} \longrightarrow$ good baseline for finding (counter) examples

2) Yes! If $q_i \neq 0$ for all $i \in \{1, \dots, n\}$ (no vectors in the set are zero) and $q_i^T q_j = 0$ if $i \neq j$ (vectors are pairwise orthogonal)

Why are orthonormal vectors useful?
$\longrightarrow$ They make projections easier (see section on QR decomposition)
(there's also other answers to this, change of basis)

Exercise If q_1 and q_2 are orthonormal vectors in $\mathbb{R}^5$, what combination $\alpha q_1 + \beta q_2$ is closest to a given vector b?

<u>Solution</u> The projection of b onto span(q_1, q_2):

$\alpha = q_1 \cdot b, \; \beta = q_2 \cdot b.$

We call $Q \in \mathbb{R}^{n \times n}$ orthogonal if $Q^T Q = I$.

Important: Q has to be a square matrix and have orthonormal columns. Orthogonal columns are not sufficient.

Properties:
1. $Q^{-1} = Q^T$
2. the columns and rows of Q form orthonormal bases of $\mathbb{R}^n$
3. they preserve lengths: $\|Qx\| = \|x\|$

1. $Q^T Q = I$ for $Q \in \mathbb{R}^{n \times n}$ implies $Q^{-1} = Q^T$ (cf. exercise 5)
2. $(Q^T Q)_{ij} = q^T_i q_j = \delta_{ij}$ where q_i for $1 \leq i \leq n$ is a column of Q. Hence the columns of Q are orthonormal per definition of an orthonormal set. We can apply the same argument with QQ^T and the rows of Q. There is no proof stated here but orthonormal sets are always linearly independent (pairwise orthogonality is not enough to follow this, all vectors have to be non zero too!). n linearly independent vectors in $\mathbb{R}^n$ form a basis.
3. $\|Qx\| = \sqrt{\langle Qx, Qx \rangle} = \sqrt{(Qx)^T Qx} = \sqrt{x^T Q^T Q x} = \sqrt{x^T x}$
$= \|x\|$

What kinds of linear transformations do orthogonal matrices correspond to?

- Rotations, reflections

With the length preserving property $\|Qx\| = \|x\|$ these two are the only options!

Exercise

Given are orthogonal matrices $A, B \in \mathbb{R}^{n \times n}$:

Which of the following is true?

a) A^T is orthogonal Proof: $A^T (A^T)^T = A^T A = I$

b) $A+B$ is orthogonal Counterexample: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

c) $A+A^T$ is orthogonal Counterexample: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

d) AB^{-1} is orthogonal

Proof:
$$(AB^{-1})(AB^{-1})^T = AB^{-1}(B^{-1})^T A^T$$
$$= AB^T B A^T = AA^T = I$$

The Gram-Schmidt Algorithm

input: $\{a_1, \dots, a_n\}$, linearly independent

output: $\{u_1, \dots, u_n\}$, orthonormal

$\operatorname{span}\{u_1, \dots, u_k\} = \operatorname{span}\{a_1, \dots, a_k\}$ for all $1 \le k \le n$

Pseudocode

$$u_1 = \frac{a_1}{\|a_1\|}$$

for $k = 2, \dots, n$

scalar product

$$\tilde{u}_k = a_k - \underbrace{\sum_{i=1}^{k-1} \langle a_k, u_i \rangle u_i}_{\text{projection of } a_k \text{ onto } \operatorname{span}(u_1, \dots, u_{k-1})}$$

$$u_k = \frac{\tilde{u}_k}{\|\tilde{u}_k\|}$$

Remember it in terms of this

Demo

https://www.desmos.com/3d/ac00d3e14b

Exercise

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{bmatrix}$$

Apply the Gram-Schmidt algorithm on the columns of A collecting them in a matrix Q. Then factorize A into A=QR.

Solution:

$$u_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\tilde{u}_2 = a_2 - \langle a_2, u_1 \rangle u_1 = \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \cdot 2 = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} \Rightarrow u_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\tilde{u}_3 = a_3 - \langle a_3, u_2 \rangle u_2 - \langle a_3, u_1 \rangle u_1 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} - 6 \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - 4 \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 5 \\ 0 \end{bmatrix} \qquad u_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 5 \end{bmatrix}$$

Assumptions:

- $A \in \mathbb{R}^{m\times n}$
- $\text{rank}(A) = n$

$$A = \underbrace{Q}_{m\times n} \; \underbrace{R}_{n\times n}$$ decomposition

Q ($m\times n$):

- output from Gram-Schmidt
- orthonormal columns (not always orthogonal)

R ($n\times n$):

- $R = Q^T b$
- upper triangular
- invertible
- $(R)_{ij} = \langle v_i, a_j \rangle$ in Gram Schmidt

What is this good for?

- Like LU decomposition for least squares

$$A^TAx = A^Tb$$

$$\iff (QR)^T QRx = (QR)^T b \qquad (A = QR)$$

$$\iff R^TQ^TQRx = R^TQ^Tb \qquad ((AB)^T = B^TA^T)$$

$$\iff R^TRx = R^TQ^Tb \qquad (Q^TQ = I)$$

$$\iff \underbrace{Rx = Q^Tb}_{\text{solve efficiently via backward substitution}} \qquad (\text{multiplying with } (R^T)^{-1} \text{ from left})$$

- Makes projections easier

$$\begin{aligned}
\text{proj}_{C(A)}(b) &= A(A^TA)^{-1}A^Tb \\
&= QR((QR)^TQR)^{-1}(QR)^Tb && (A = QR) \\
&= QR(R^TQ^TQR)^{-1}R^TQ^Tb && ((AB)^T = B^TA^T) \\
&= QR(R^TR)^{-1}R^TQ^Tb && (Q^TQ = I) \\
&= QRR^{-1}(R^T)^{-1}R^TQ^Tb && ((AB)^{-1} = B^{-1}A^{-1}) \\
&= QQ^Tb
\end{aligned}$$

Exercise Prove $Qx = 0 \Rightarrow x = 0$ when Q has orthogonal columns without saying the word linear independence.

<u>Solution:</u> We multiply with Q^T from the left:

$$Qx = 0 \iff Q^T Q x = Q^T 0 = 0$$
$$\iff x = 0$$

Pseudoinverses

If there's no inverse, can we find something as close to it as possible? $\longrightarrow A^+$

Let $A \in \mathbb{R}^{m \times n}$:

left inverse if <u>full column rank</u>: $A^+ A = I$

$$A^+ = \underbrace{(A^T A)}_{\text{invertible } n \times n \text{ matrix if rank}(A) = n}{}^{-1} A^T \qquad (1)$$

right inverse if <u>full row rank</u>: $A A^+ = I$

$$A^+ = A^T \underbrace{(A A^T)}_{\text{invertible } m \times m \text{ matrix if rank}(A) = m}{}^{-1} \qquad (2)$$

<u>General case: rank$(A) = r$</u>

$$A^+ = R^+ C^+ \qquad A = \underbrace{C}_{\text{first independent columns}} \underbrace{R}_{\text{rref}(A) \text{ without zero rows}}, \quad C \in \mathbb{R}^{m \times r}, \quad R \in \mathbb{R}^{r \times n}$$

(holds for any full rank decomposition)

$$A^+ = R^+ C^+$$
$$= R^T (R R^T)^{-1} (C^T C)^{-1} C^T \quad ((1),(2) \text{ for } C, R)$$
$$= R^T (C^T C R R^T)^{-1} C^T \quad ((AB)^{-1} = B^{-1} A^{-1})$$
$$= R^T (C^T A R^T)^{-1} C^T \quad (A = CR)$$

Examples

$$A_1 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad A_2 = \begin{bmatrix} 2 & 0 \\ 0 & 4 \\ 0 & 0 \end{bmatrix} \quad A_3 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \end{bmatrix}$$

We apply the formula $A^+ = R^T (C^T A R^T)^{-1} C^T$ and get:

$$A_1^+ = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \frac{1}{4} \begin{bmatrix} 1 & 1 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad A_2^+ = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/4 & 0 \end{bmatrix}$$

$$A_3^+ = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/4 \\ 0 & 0 \end{bmatrix}$$

References:
Last years course
https://github.com/mitmath/1806