Discussion 1

1. Eigenvalues

Let $A \in \mathbb{R}^{n\times n}$ have the eigendecomposition $P\Lambda P^{-1}$ where $\Lambda \in \mathbb{R}^{n\times n}$ is a diagonal matrix with entries consisting of the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ and $P \in \mathbb{R}^{n\times n}$ is an invertible matrix. Note that this is equivalent to stating that A is diagonalizable via the transformation,

$$P^{-1}AP = \Lambda. \tag{1}$$

(a) Show that $A^m = P\Lambda^m P^{-1}$, for integer $m \geq 1$.

Solution:

$$\begin{aligned} A^m &= (P\Lambda P^{-1})(P\Lambda P^{-1})\dots(P\Lambda P^{-1}) \quad m \text{ times} && (2)\\ &= P\Lambda(P^{-1}P)\Lambda(P^{-1}P)\dots\Lambda(P^{-1}P)\Lambda P^{-1} && (3)\\ &= P\Lambda^m P^{-1}. && (4) \end{aligned}$$

The last equality follows from the repeated application of the identity $P^{-1}P = I$.

(b) Show that determinant of A is the product of its eigenvalues, i.e.

$$\det(A) = \prod_{i=1}^{n} \lambda_i. \tag{5}$$

HINT: We have the identity $\det(XY) = \det(X)\det(Y)$.

Solution: Write down eigendecomposition of A and use properties of determinant given in the hint.

$$\begin{aligned} \det(A) &= \det\left(P\Lambda P^{-1}\right) && (6)\\ &= \det(P)\det(\Lambda)\det\left(P^{-1}\right) && (7)\\ &= \det\left(PP^{-1}\right)\det(\Lambda) && (8)\\ &= \det(\Lambda) && (9)\\ &= \prod_{i=1}^{n} \lambda_i && (10) \end{aligned}$$

2. Least Squares and Gram-Schmidt

Consider the least squares problem

$$\vec{x}^\star = \underset{\vec{x}\in\mathbb{R}^n}{\operatorname{argmin}} \left\|A\vec{x} - \vec{b}\right\|_2^2 \tag{11}$$

where $A \in \mathbb{R}^{m\times n}$, $\vec{b} \in \mathbb{R}^m$ and assume A is full column rank. One way to solve this least-squares problem is to use Gram-Schmidt Orthonormalization (GSO). Using GSO, the matrix A can be written as,

$$A = QR = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \tag{12}$$

where Q is an orthonormal matrix and R is an upper-triangular matrix. The columns of Q_1 form an orthonormal basis for the range space $\mathcal{R}(A)$ and columns of Q_2 form an orthonormal basis for $\mathcal{R}(A)^\perp$. Moreover, R_1 is upper triangular and invertible.

(a) Show that the squared norm of the residual is given by

$$\|\vec{r}\|_2^2 := \left\|\vec{b} - A\vec{x}\right\|_2^2 = \left\|Q_1^\top \vec{b} - R_1\vec{x}\right\|_2^2 + \left\|Q_2^\top \vec{b}\right\|_2^2. \tag{13}$$

Solution: We have,

$$\|\vec{r}\|_2^2 := \left\|\vec{b} - A\vec{x}\right\|_2^2 \tag{14}$$

$$= \left\|\vec{b} - Q\begin{bmatrix} R_1 \\ 0 \end{bmatrix}\vec{x}\right\|_2^2. \tag{15}$$

Since multiplying by an orthonormal matrix does not change the ℓ_2-norm of a vector we can multiply by $Q^\top$ to get,

$$\|\vec{r}\|_2^2 = \left\|Q^\top\left(\vec{b} - Q\begin{bmatrix} R_1 \\ 0 \end{bmatrix}\vec{x}\right)\right\|_2^2 \tag{16}$$

$$= \left\|\begin{bmatrix} Q_1^\top \vec{b} \\ Q_2^\top \vec{b} \end{bmatrix} - \begin{bmatrix} R_1\vec{x} \\ 0 \end{bmatrix}\right\|_2^2 \tag{17}$$

$$= \left\|\begin{bmatrix} Q_1^\top \vec{b} - R_1\vec{x} \\ Q_2^\top \vec{b} \end{bmatrix}\right\|_2^2 \tag{18}$$

$$= \left\|Q_1^\top \vec{b} - R_1\vec{x}\right\|_2^2 + \left\|Q_2^\top \vec{b}\right\|_2^2. \tag{19}$$

(b) Find $\vec{x}^\star$ such that the squared norm of the residual in (13) is minimized. Your expression for $\vec{x}^\star$ should only use some or all of the following terms: Q_1, Q_2, R_1, $\vec{b}$.

Solution: We have,

$$\|\vec{r}\|_2^2 = \left\|Q_1^\top \vec{b} - R_1\vec{x}\right\|_2^2 + \left\|Q_2^\top \vec{b}\right\|_2^2. \tag{20}$$

Since we have no control over the term $\left\|Q_2^\top \vec{b}\right\|_2^2$ (i.e., no matter how we change $\vec{x}$, that term stays constant because it doesn't involve $\vec{x}$ at all), it is irrelevant from the perspective of the optimization, and so the optimal $\vec{x}^\star$ is one which minimizes $\left\|Q_1^\top \vec{b} - R_1\vec{x}\right\|_2^2$. This expression is minimized when $Q_1^\top \vec{b} = R_1\vec{x}$, and using the fact that R_1 is invertible we have $\vec{x}^\star = R_1^{-1}Q_1^\top \vec{b}$.

(c) Check if the expression for $\vec{x}^{\star}$ obtained in the previous part is equivalent to the one obtained by the formula, $\vec{x}^{\star} = (A^{\top}A)^{-1}A^{\top}\vec{b}$.

Solution: We have $A = QR = Q_1 R_1$ (block multiplication for matrices). Substituting,

$$\begin{aligned} \vec{x}^{\star} &= (R_1^{\top} R_1)^{-1} R_1^{\top} Q_1^{\top} \vec{b} && (21) \\ &= R_1^{-1} (R_1^{\top})^{-1} R_1^{\top} Q_1^{\top} \vec{b} && (22) \\ &= R_1^{-1} Q_1^{\top} \vec{b}. && (23) \end{aligned}$$

This is the same as we got in the previous part.

3. Invertibility of $A^\top A$

In this problem, we show that if the matrix $A \in \mathbb{R}^{m\times n}$ has a full column rank, then the matrix $A^\top A$ is invertible.

(a) Show that if a vector $\vec{x}$ is in the null space of A then $\vec{x}$ is in the null space of $A^\top A$.

Solution:

$$\begin{aligned} \vec{x} \in \mathcal{N}(A) &\iff A\vec{x} = \vec{0} && (24) \\ &\implies A^\top A\vec{x} = \vec{0} && (25) \\ &\iff \vec{x} \in \mathcal{N}\left(A^\top A\right) && (26) \end{aligned}$$

Where line 25 follows by multiplying both sides of $A\vec{x} = 0$ by $A^\top$

(b) Conversely, show that if $\vec{x}$ is in the null space of $A^\top A$ then $\vec{x}$ is in the null space of A.

Solution:

$$\begin{aligned} \vec{x} \in \mathcal{N}\left(A^\top A\right) &\iff A^\top A\vec{x} = \vec{0} && (27) \\ &\implies \vec{x}^\top A^\top A\vec{x} = \vec{0} && (28) \\ &\implies (A\vec{x})^\top A\vec{x} = \vec{0} && (29) \\ &\implies \|A\vec{x}\|_2^2 = 0 && (30) \\ &\implies A\vec{x} = 0 && (31) \\ &\implies \vec{x} \in \mathcal{N}(A) && (32) \end{aligned}$$

Where line 28 follows by multiplying both sides of $A^\top A\vec{x} = \vec{0}$ by $\vec{x}^\top$ and line 31 follows from the properties of norms.

(c) Given that matrix A has a full column rank, what can you say about its null space? What does this imply about the null space and invertibility of the matrix $A^\top A$?

Solution: $\mathcal{N}(A) = \{\vec{0}\}$. From the previous parts, we have shown that $\mathcal{N}(A) = \mathcal{N}\left(A^\top A\right)$ then $\mathcal{N}\left(A^\top A\right) = \{\vec{0}\}$ and thus $A^\top A$ is invertible.