

1. Least Squares and Gram-Schmidt

Consider the least squares problem

$$\vec{x}^* = \underset{\vec{x} \in \mathbb{R}^n}{\operatorname{argmin}} \|A\vec{x} - \vec{b}\|_2^2 \quad (1)$$

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,

$$A = QR = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \quad (2)$$

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 the range space $\mathcal{R}(A)^\perp$. Moreover, R_1 is upper triangular and invertible.

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

$$\|\vec{r}\|_2^2 := \|\vec{b} - A\vec{x}\|_2^2 = \|Q_1^\top \vec{b} - R_1 \vec{x}\|_2^2 + \|Q_2^\top \vec{b}\|_2^2. \quad (3)$$

Solution: We have,

$$\|\vec{r}\|_2^2 := \|\vec{b} - A\vec{x}\|_2^2 \quad (4)$$

$$= \left\| \vec{b} - Q \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \vec{x} \right\|_2^2. \quad (5)$$

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

$$\|\vec{r}\|_2^2 = \left\| Q^\top \left(\vec{b} - Q \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \vec{x} \right) \right\|_2^2 \quad (6)$$

$$= \left\| \begin{bmatrix} Q_1^\top \vec{b} \\ Q_2^\top \vec{b} \end{bmatrix} - \begin{bmatrix} R_1 \vec{x} \\ 0 \end{bmatrix} \right\|_2^2 \quad (7)$$

$$= \left\| \begin{bmatrix} Q_1^\top \vec{b} - R_1 \vec{x} \\ Q_2^\top \vec{b} \end{bmatrix} \right\|_2^2 \quad (8)$$

$$= \|Q_1^\top \vec{b} - R_1 \vec{x}\|_2^2 + \|Q_2^\top \vec{b}\|_2^2. \quad (9)$$

- (b) Find \vec{x}^* such that the squared norm of the residual in Equation (3) is minimized. Your expression for \vec{x}^* should only use some or all of the following terms: Q_1, Q_2, R_1, \vec{b} .

Solution: We have,

$$\|\vec{r}\|_2^2 = \|Q_1^\top \vec{b} - R_1 \vec{x}\|_2^2 + \|Q_2^\top \vec{b}\|_2^2. \quad (10)$$

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}^* 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 $\vec{x}^* = R_1^{-1} Q_1^\top \vec{b}$.

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

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

$$\vec{x}^* = (R_1^\top R_1)^{-1} R_1^\top Q_1^\top \vec{b} \quad (11)$$

$$= R_1^{-1} (R_1^\top)^{-1} R_1^\top Q_1^\top \vec{b} \quad (12)$$

$$= R_1^{-1} Q_1^\top \vec{b}. \quad (13)$$

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

2. Eigenvalues

- (a) Let $A \in \mathbb{R}^{n \times n}$ be a matrix with eigenvalues and corresponding eigenvectors $\lambda_1, \dots, \lambda_n$ and $\vec{v}_1, \dots, \vec{v}_n$ respectively. Now consider $B = A + cI_n$ where $c \in \mathbb{R}$ and I_n is the $n \times n$ identity matrix. What are the eigenvalues and eigenvectors of B in terms of c and λ_i, \vec{v}_i for $i = 1, \dots, n$?

Solution: Suppose λ, \vec{v} is an eigenpair of A . Then

$$B\vec{v} = (A + cI)\vec{v} \quad (14)$$

$$= A\vec{v} + cI\vec{v} \quad (15)$$

$$= \lambda\vec{v} + c\vec{v} \quad (16)$$

$$= (\lambda + c)\vec{v}. \quad (17)$$

That is, $A + cI$ has the same eigenvectors as A , and the eigenvalues are just shifted by c .

- (b) Let Q be an orthonormal matrix, i.e., $Q^\top Q = I$ of size $n \times n$. Let $\lambda \in \mathbb{R}$ be a scalar, $\vec{v} \in \mathbb{R}^n$ be a vector, and $A \in \mathbb{R}^{n \times n}$ be a matrix. Prove that if we have

$$A\vec{v} = \lambda\vec{v}, \quad (18)$$

i.e., \vec{v}, λ is an eigenpair of A , then we have

$$(QAQ^\top)(Q\vec{v}) = \lambda(Q\vec{v}). \quad (19)$$

Solution: As $A\vec{v} = \lambda\vec{v}$ and $Q^\top Q = I$, we have

$$AQ^\top Q\vec{v} = \lambda\vec{v}. \quad (20)$$

Multiplying both sides with Q we get

$$(QAQ^\top)(Q\vec{v}) = \lambda(Q\vec{v}) \quad (21)$$

as desired.

- (c) Let A be a $d \times n$ matrix. Prove that the non-zero eigenvalues of AA^\top are the same as the non-zero eigenvalues of $A^\top A$.

Solution: Say $\lambda \neq 0, \vec{v}$ is an eigenpair of $A^\top A$ which is a $n \times n$ matrix. Hence,

$$(A^\top A)\vec{v} = \lambda\vec{v} \quad (22)$$

Multiply both sides with A , to get,

$$A(A^\top A)\vec{v} = A(\lambda\vec{v}) \quad (23)$$

$$(AA^\top)A\vec{v} = \lambda A\vec{v} \quad (24)$$

As $\lambda \neq 0$ and $\vec{v} \neq 0_n$, we have $\lambda\vec{v} \neq 0_n$ and so, $(A^\top A)\vec{v} \neq 0_n$. Thus $A^\top(A\vec{v}) \neq 0_n$, which implies that $A\vec{v} \neq 0_d$. Therefore, $A\vec{v}$ is an eigenvector of AA^\top corresponding to λ . Similarly, we can show that every non-zero eigenvalue of AA^\top is an eigenvalue of $A^\top A$ and we are done.

- (d) Given a matrix $A = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$ find its eigenvalues and eigenvectors without using the characteristic polynomial.

HINT: Use the fact that eigenvectors with eigenvalue 0 span the null space. Also notice that the eigenvectors of a symmetric matrix corresponding to different eigenvalues are orthogonal to each other.

Solution: First notice that the matrix is of rank 1 because both columns are the same. Hence, one of the eigenvalues has to be $\lambda_2 = 0$. Its eigenvector is just the basis of the null space which is

$$\vec{v}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}. \quad (25)$$

Here are two methods to compute the second eigenpair:

Approach 1:

We can first observe that the matrix is symmetric and hence the set of its eigenvectors form an orthonormal basis. We know three things:

- i. We know the value of \vec{v}_2 .
- ii. We know that the whole space is \mathbb{R}^2 .
- iii. We know that \vec{v}_1 is orthogonal to \vec{v}_2 .

From these three facts, we can look for vectors orthogonal to \vec{v}_2 in \mathbb{R}^2 . This can be done in many ways (graphically or via Gram-Schmidt or similar), and in the end we get

$$\vec{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}. \quad (26)$$

Now by brute-force computation, we see $A\vec{v}_1 = 4\vec{v}_1$ and hence $\lambda_1 = 4$.

Approach 2:

Another way to compute λ_1, \vec{v}_1 , once we have computed λ_2, \vec{v}_2 is via the definition of the Frobenius norm. We know that for any symmetric matrix, $\|A\|_F^2 = \sum_{i=1}^n \lambda_i^2$. That is, the squared Frobenius norm is the sum of its eigenvalues. But another definition of $\|A\|_F^2 = \sum_{i=1}^n \sum_{j=1}^n A_{i,j}^2$. That is, the squared Frobenius norm is the sum of squares of all the entries of the matrix. Using the first definition, $\|A\|_F^2 = \lambda_1^2 + \lambda_2^2 = \lambda_1^2$. Using the second definition, $\|A\|_F^2 = 16$. So $\lambda_1^2 \in \{-4, 4\}$. Finally, we can show that A is positive semidefinite by showing that $\vec{x}^\top A \vec{x} \geq 0$ for all \vec{x} . Thus all its eigenvalues are non-negative. So $\lambda_1 = 4$. The eigenvector can be computed as a basis vector of the null space of $A - 4I$.

The reason to not use the characteristic polynomial method for certain matrices is that computing determinants can be expensive in general. But when we have matrices with more structure, we can simplify the problem.

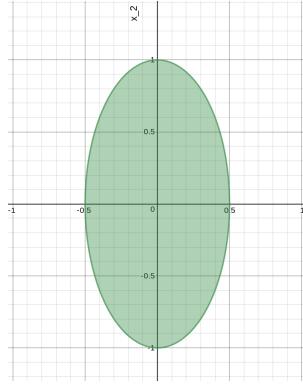


Figure 1: Region corresponding to $4x_1^2 + x_2^2 \leq 1$.

3. Symmetric Matrices

- (a) Show that any symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive semidefinite if and only if there exists a symmetric matrix $C \in \mathbb{R}^{n \times n}$ such that $A = C^\top C$.

Solution: Forward direction:

$$\vec{x}^\top C^\top C \vec{x} = \|C\vec{x}\|_2^2 \geq 0 \quad (27)$$

and hence $C^\top C$ is PSD. Reverse direction: If A is PSD, then $A = UDU^\top$ by the spectral theorem with all the diagonal elements of D non-negative. Let $D^{1/2}$ be the square root of D (in this case just the square root of all the diagonal elements) and set $C = UD^{1/2}U^\top$. Observe that C is symmetric. Note that there are other non-symmetric C which also satisfy $C^\top C = A$. For example, $C = D^{1/2}U^\top$.

- (b) Draw the region $\left\{ \vec{x} \in \mathbb{R}^2 \mid \vec{x}^\top \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \vec{x} \leq 1 \right\}$.

Solution: Simplifying we get

$$4x_1^2 + x_2^2 \leq 1. \quad (28)$$

The corresponding region is an ellipse with minor axis 1 and major-axis 2 as shown in Fig. 1:

- (c) Draw the region $\left\{ \vec{x} \in \mathbb{R}^2 \mid \vec{x}^\top \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \vec{x} \leq 1 \right\}$.

Solution: Call the given matrix as A . The eigenvalues of A can be found by solving,

$$\det(A - \lambda I) = 0 \quad (29)$$

$$\implies (1 - \lambda)^2 - 1 = 0 \quad (30)$$

$$\implies \lambda = 0, \text{ or } \lambda = 2. \quad (31)$$

The eigenvector associated with eigenvalue 0, say \vec{v}_1 , is $\left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right]^\top$. The eigenvector associated with eigenvalue 2, say \vec{v}_2 , is $\left[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right]$. Observe that \vec{v}_1, \vec{v}_2 are orthonormal vectors, and any $\vec{x} \in \mathbb{R}^2$ can be written as $\vec{x} = a\vec{v}_1 + b\vec{v}_2$, for $a, b \in \mathbb{R}$. Then,

$$\vec{x}^\top A \vec{x} = (a\vec{v}_1 + b\vec{v}_2)^\top A (a\vec{v}_1 + b\vec{v}_2) \quad (32)$$

$$= a^2 \vec{v}_1^\top A \vec{v}_1 + b^2 \vec{v}_2^\top A \vec{v}_2 + ab \vec{v}_1^\top A \vec{v}_2 + ba \vec{v}_2^\top A \vec{v}_1 \quad (33)$$

$$= 2b^2. \quad (34)$$

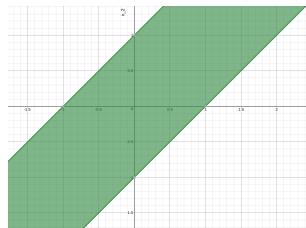


Figure 2: Region corresponding to $x_1^2 + x_2^2 - 2x_1x_2 \leq 1$.

Thus we have the condition $|b| \leq \frac{1}{\sqrt{2}}$. Note that a is a free parameter and can take any value in \mathbb{R} . Thus the region is a strip parallel to the line $y = x$ of width $\sqrt{2}$ as shown in Fig. 2. Note that the strip is not bounded and any $\vec{x} = a\vec{v}_1$ satisfies the condition for all $a \in \mathbb{R}$.

- (d) Why is the region in part (b) bounded, whereas the region in part (c) is unbounded?

Solution: This is because A in part (b) is positive definite while the A in part (c) is positive semi-definite and has a non-trivial nullspace. Suppose \vec{v} is a unit vector in the nullspace of A . Then $\vec{x} = t\vec{v}$ will satisfy the equation for all $t \in \mathbb{R}$, however large.