Discussion 2: Least Squares, Eigenvalues, and Symmetric Matrices

1. Review: Key Concepts

Before diving into the problems, we review the essential definitions and properties.

1.1 QR Factorization

Definition 1 (QR Factorization)**.** Any matrix $A \in \mathbb{R}^{m\times n}$ with full column rank can be factored as:

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

where $Q \in \mathbb{R}^{m\times m}$ is orthonormal ($Q^\top Q = I$), $R_1 \in \mathbb{R}^{n\times n}$ is upper triangular and invertible, the columns of Q_1 form an orthonormal basis for $\mathcal{R}(A)$, and the columns of Q_2 form an orthonormal basis for $\mathcal{R}(A)^\perp$.

Key property: Multiplying by an orthonormal matrix preserves norms: $\|Q^\top x\|_2 = \|x\|_2$ for any orthonormal Q.

1.2 Least Squares

Definition 2 (Least Squares Problem)**.** Given $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, the least squares solution minimizes the residual:

$$x^\star = \arg\min_{x\in\mathbb{R}^n} \|Ax - b\|_2^2$$

Normal Equations: When A has full column rank, the unique solution is:

$$x^\star = (A^\top A)^{-1} A^\top b$$

1.3 Eigenvalues and Eigenvectors

Recall that (λ, v) is an eigenpair of A if $Av = \lambda v$ with $v \neq 0$. For a symmetric matrix $A = A^\top$, eigenvectors corresponding to distinct eigenvalues are orthogonal.

1.4 Positive Semidefinite Matrices

Definition 3 (PSD and PD)**.** A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is **positive semidefinite** (PSD) if $x^\top Ax \geq 0$ for all $x \in \mathbb{R}^n$, and **positive definite** (PD) if $x^\top Ax > 0$ for all $x \neq 0$.

Spectral characterization: A symmetric matrix is PSD $\iff$ all eigenvalues ≥ 0, and PD $\iff$ all eigenvalues > 0.

2. Problem 1: Least Squares and Gram-Schmidt

Consider the least squares problem $x^\star = \arg\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2$ where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and A has full column rank. Using the QR factorization $A = QR = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}\begin{bmatrix} R_1 \\ 0 \end{bmatrix}$.

2.1 Part (a): Residual Norm Decomposition

Goal: Show that $\|b-Ax\|_2^2 = \|Q_1^\top b - R_1 x\|_2^2 + \|Q_2^\top b\|_2^2$.

Step 1: Substitute the QR factorization.

Since $A = Q\begin{bmatrix} R_1 \\ 0 \end{bmatrix}$, we write:

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

Step 2: Multiply by $Q^\top$ (preserves norms).

Since Q is orthonormal, $\|Q^\top y\|_2 = \|y\|_2$ for any y. Multiplying the argument by $Q^\top$:

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

Step 3: Expand the block structure.

Using $Q^\top b = \begin{bmatrix} Q_1^\top b \\ Q_2^\top b \end{bmatrix}$:

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

Step 4: Split the squared norm.

The squared norm of a stacked vector equals the sum of the squared norms:

$$= \|Q_1^\top b - R_1 x\|_2^2 + \|Q_2^\top b\|_2^2 \quad \square$$

Intuition

The QR factorization decomposes the residual into two independent pieces: one that depends on x (the projection onto $\mathcal{R}(A)$) and one that is fixed regardless of x (the component in $\mathcal{R}(A)^\perp$).

2.2 Part (b): Optimal $x^\star$ via QR

Goal: Find $x^\star$ minimizing $\|b-Ax\|_2^2$ using the decomposition from part (a).

Step 1: Identify what we can control.

From part (a):

$$\|b-Ax\|_2^2 = \underbrace{\|Q_1^\top b - R_1 x\|_2^2}_{\text{depends on } x} + \underbrace{\|Q_2^\top b\|_2^2}_{\text{constant w.r.t. } x}$$

The second term does not involve x, so minimizing over x only affects the first term.

Step 2: Set the controllable term to zero.

The minimum of $\|Q_1^\top b - R_1 x\|_2^2$ is 0, achieved when $R_1 x = Q_1^\top b$.

Step 3: Solve for $x^\star$.

Since R_1 is upper triangular and invertible:

$$x^\star = R_1^{-1} Q_1^\top b$$

2.3 Part (c): Equivalence with Normal Equations

Goal: Verify that $R_1^{-1}Q_1^\top b = (A^\top A)^{-1}A^\top b$.

Step 1: Express A in terms of Q_1 and R_1.

By block multiplication: $A = QR = Q_1 R_1$ (since the Q_2 block multiplies the zero block).

Step 2: Substitute into the normal equations formula.

$$\begin{aligned}(A^\top A)^{-1}A^\top b &= \big((Q_1R_1)^\top (Q_1R_1)\big)^{-1}(Q_1R_1)^\top b \\ &= (R_1^\top Q_1^\top Q_1 R_1)^{-1} R_1^\top Q_1^\top b \\ &= (R_1^\top R_1)^{-1} R_1^\top Q_1^\top b\end{aligned}$$

Step 3: Simplify using the inverse of a product.

$$(R_1^\top R_1)^{-1}R_1^\top = R_1^{-1}(R_1^\top)^{-1}R_1^\top = R_1^{-1}$$

Step 4: Conclude.

$$(A^\top A)^{-1}A^\top b = R_1^{-1}Q_1^\top b = x^\star \quad \square$$

Why QR is Better in Practice

While both formulas give the same answer, the QR approach ($x^\star = R_1^{-1}Q_1^\top b$) is numerically more stable than forming $A^\top A$ and solving the normal equations. Computing $A^\top A$ squares the condition number, amplifying numerical errors.

3. Problem 2: Eigenvalues

3.1 Part (a): Eigenvalues of $A + cI$

Goal: Given that A has eigenpairs (λ_i, v_i), find the eigenpairs of $B = A + cI$.

Step 1: Apply B to an eigenvector of A.

Suppose (λ, v) is an eigenpair of A, so $Av = \lambda v$. Then:

$$\begin{aligned} Bv &= (A + cI)v \\ &= Av + cIv \\ &= \lambda v + cv \\ &= (\lambda + c)v \end{aligned}$$

Step 2: Conclude.

$B = A + cI$ has the **same eigenvectors** as A, and the eigenvalues are **shifted by** c: eigenvalue λ_i becomes $\lambda_i + c$.

Eigenvalue Shifting

Adding cI to a matrix shifts the entire spectrum by c without changing the eigenvectors. This is useful for making matrices invertible (shift away from zero eigenvalues) or for making matrices PSD (shift eigenvalues to be non-negative).

3.2 Part (b): Eigenpairs of $QAQ^\top$

Goal: Given $Av = \lambda v$ and $Q^\top Q = I$, show that $(QAQ^\top)(Qv) = \lambda(Qv)$.

Step 1: Insert $Q^\top Q = I$ into the eigenvector equation.

Since $Q^\top Q = I$, we can write:

$$Av = A(Q^\top Q)v = \lambda v$$

Step 2: Multiply both sides by Q.

$$\begin{aligned} QA(Q^\top Q)v &= \lambda Qv \\ (QAQ^\top)(Qv) &= \lambda(Qv) \quad \square \end{aligned}$$

If (λ, v) is an eigenpair of A, then (λ, Qv) is an eigenpair of $QAQ^\top$. Orthogonal similarity preserves eigenvalues and transforms eigenvectors by Q.

3.3 Part (c): Non-zero Eigenvalues of $AA^\top$ and $A^\top A$

Goal: For $A \in \mathbb{R}^{d\times n}$, prove that the non-zero eigenvalues of $AA^\top$ equal those of $A^\top A$.

Step 1: Start with an eigenpair of $A^\top A$.

Suppose $\lambda \neq 0$ and v is an eigenvector of $A^\top A$:

$$(A^\top A)v = \lambda v$$

Step 2: Multiply both sides by A.

$$\begin{aligned} A(A^\top A)v &= A(\lambda v) \\ (AA^\top)(Av) &= \lambda(Av) \end{aligned}$$

Step 3: Verify $Av \neq 0$.

Since $\lambda \neq 0$ and $v \neq 0$, we have $(A^\top A)v = \lambda v \neq 0$, so $A^\top(Av) \neq 0$, which implies $Av \neq 0$.

Step 4: Conclude.

Therefore Av is an eigenvector of $AA^\top$ with eigenvalue λ. The same argument in reverse (starting from an eigenpair of $AA^\top$ and multiplying by $A^\top$) shows the converse. $\square$

The non-zero eigenvalues of $AA^\top$ and $A^\top A$ are identical. (They may differ in their zero eigenvalues since the matrices have different sizes: $AA^\top \in \mathbb{R}^{d\times d}$ vs. $A^\top A \in \mathbb{R}^{n\times n}$.)

3.4 Part (d): Eigenvalues of $\left[\begin{smallmatrix} 2 & 2 \\ 2 & 2 \end{smallmatrix}\right]$ Without Characteristic Polynomial

Goal: Find eigenvalues and eigenvectors of $A = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$ without using the characteristic polynomial.

Step 1: Observe the rank.

Both columns of A are identical, so $\text{rank}(A) = 1$. This means one eigenvalue must be $\lambda_2 = 0$ (since a rank-1 matrix in $\mathbb{R}^{2\times 2}$ has a one-dimensional null space).

Step 2: Find the eigenvector for $\lambda_2 = 0$.

The eigenvector with eigenvalue 0 spans $\mathcal{N}(A)$. Solving $Ax = 0$:

$$v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Step 3: Find the other eigenvector using symmetry.

Since A is symmetric, its eigenvectors are orthogonal. A unit vector orthogonal to v_2 in $\mathbb{R}^2$ is:

$$v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Step 4: Find the corresponding eigenvalue.

Compute Av_1 directly:

$$Av_1 = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 4 \\ 4 \end{bmatrix} = 4 \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 4v_1$$

So $\lambda_1 = 4$.

Eigenvalues: $\lambda_1 = 4$, $\lambda_2 = 0$. **Eigenvectors:** $v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

Alternative: Frobenius Norm

Since A is symmetric, $\|A\|_F^2 = \sum \lambda_i^2$. Computing entry-wise: $\|A\|_F^2 = 4 \cdot 4 = 16$. With $\lambda_2 = 0$, we get $\lambda_1^2 = 16$, so $\lambda_1 = \pm 4$. Since A is PSD (verify $x^\top Ax = 2(x_1 + x_2)^2 \geq 0$), we conclude $\lambda_1 = 4$.

4. Problem 3: Symmetric Matrices

4.1 Part (a): PSD $\iff A = C^\top C$ for Symmetric C

Goal: Show that a symmetric A is PSD if and only if $A = C^\top C$ for some symmetric C.

Forward direction ($\Leftarrow$): Suppose $A = C^\top C$. For any $x \in \mathbb{R}^n$:

$$x^\top A x = x^\top C^\top C x = \|Cx\|_2^2 \geq 0$$

So A is PSD.

Reverse direction ($\Rightarrow$): Suppose A is symmetric and PSD. By the spectral theorem:

$$A = UDU^\top$$

where U is orthogonal and $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$ with all $\lambda_i \geq 0$.

Define $D^{1/2} = \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$ and set $C = UD^{1/2}U^\top$. Then:

$$C^\top C = (UD^{1/2}U^\top)^\top (UD^{1/2}U^\top) = UD^{1/2}U^\top UD^{1/2}U^\top = UDU^\top = A$$

Moreover, $C = UD^{1/2}U^\top$ is symmetric since $(UD^{1/2}U^\top)^\top = U(D^{1/2})^\top U^\top = UD^{1/2}U^\top$. $\square$

Remark

The symmetric square root $C = UD^{1/2}U^\top$ is not the only valid choice. For example, $C = D^{1/2}U^\top$ also satisfies $C^\top C = A$, but this C is generally not symmetric.

4.2 Part (b): Ellipse Region

Goal: Draw the region $\{x \in \mathbb{R}^2 \mid x^\top \left[\begin{smallmatrix} 4 & 0 \\ 0 & 1 \end{smallmatrix}\right] x \leq 1\}$.

Step 1: Expand the quadratic form.

$$x^\top \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} x = 4x_1^2 + x_2^2$$

Step 2: Identify the region.

The inequality $4x_1^2 + x_2^2 \leq 1$ can be rewritten as:

$$\frac{x_1^2}{(1/2)^2} + \frac{x_2^2}{1^2} \leq 1$$

This is an **ellipse** centered at the origin with semi-axis $\frac{1}{2}$ along x_1 and semi-axis 1 along x_2. The region is **bounded**.

4.3 Part (c): Strip Region

Goal: Draw the region $\{x \in \mathbb{R}^2 \mid x^\top \left[\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right] x \leq 1\}$.

Step 1: Expand the quadratic form.

$$x^\top \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} x = x_1^2 + x_2^2 - 2x_1x_2 = (x_1 - x_2)^2$$

Step 2: Find the eigendecomposition.

The matrix $A = \left[\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right]$ has eigenvalues $\lambda = 0$ and $\lambda = 2$, with eigenvectors:

$$v_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \ (\lambda = 0), \qquad v_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix} \ (\lambda = 2)$$

Step 3: Change coordinates.

Writing $x = av_1 + bv_2$, the quadratic form becomes $x^\top A x = 2b^2$.

The constraint $2b^2 \leq 1$ gives $|b| \leq \frac{1}{\sqrt{2}}$, with a free (any real value).

The region is an **unbounded strip** of width $\sqrt{2}$ centered on the line $x_1 = x_2$ (the direction of eigenvector v_1). Any point $x = av_1$ along this line satisfies the constraint for all $a \in \mathbb{R}$.

4.4 Part (d): Bounded vs. Unbounded Regions

Goal: Explain why part (b) is bounded and part (c) is unbounded.

The matrix in part (b) has eigenvalues 4 and 1, both strictly positive, so it is **positive definite**. For a PD matrix, $x^\top Ax > 0$ for all $x \neq 0$, which forces $\|x\|$ to be bounded.

The matrix in part (c) has eigenvalues 2 and 0, so it is **positive semidefinite** (but not PD). The zero eigenvalue means there exists a direction v (the null space) along which $x^\top Ax = 0$ regardless of how far we go. Specifically, $x = tv_1$ satisfies the constraint for all $t \in \mathbb{R}$.

PD matrix $\Rightarrow$ bounded sublevel set. **PSD (not PD) matrix $\Rightarrow$ unbounded sublevel set.**
The null space of a PSD matrix provides directions along which the quadratic form vanishes, allowing the region to extend infinitely.

Key Takeaways

1. **QR and least squares:** The QR factorization decomposes the residual into a controllable part and a fixed part. The optimal solution is $x^\star = R_1^{-1}Q_1^\top b$, equivalent to the normal equations but numerically more stable.
2. **Eigenvalue shifting:** Adding cI shifts all eigenvalues by c without changing eigenvectors. Orthogonal similarity $QAQ^\top$ preserves eigenvalues and transforms eigenvectors by Q.
3. **Shared non-zero eigenvalues:** $AA^\top$ and $A^\top A$ have the same non-zero eigenvalues. The trick is to multiply the eigenvector equation by A (or $A^\top$) to transfer eigenpairs between the two.
4. **Structured eigenvalue computation:** For structured matrices (e.g., rank-1), use null space analysis and orthogonality of symmetric eigenvectors rather than the characteristic polynomial.
5. **PSD geometry:** PD matrices yield bounded ellipsoids; PSD matrices with zero eigenvalues yield unbounded strips. The null space determines the unbounded directions.