

- 1** (a) Use the Gram-Schmidt process to transform the basis $\mathbf{w}_1 = (1, 2, 3)$, $\mathbf{w}_2 = (3, 2, 7)$ into an orthonormal basis.
- (b) Find the orthogonal projection of $\mathbf{x} = (1, 1, -2)$ onto the subspace of R^3 spanned by the orthonormal basis in (a).
- (c) Find a QR decomposition of $A = [\mathbf{w}_1 \ \mathbf{w}_2]$.
- (d) Find a least squares solution of $A\mathbf{x} = \mathbf{b}$ using the QR decomposition in (c),
where $\mathbf{b} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$.

Solution.

- (a) Let $\mathbf{v}_1 = \mathbf{w}_1$. Then, by using the Gram-Schmidt process we have

$$\mathbf{v}_2 = \mathbf{w}_2 - \frac{\mathbf{v}_1 \cdot \mathbf{w}_2}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 = (3, 2, 7) - \frac{28}{14}(1, 2, 3) = (1, -2, 1). \quad (+6 \text{ points})$$

$\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for the subspace of R^3 spanned by $\{\mathbf{w}_1, \mathbf{w}_2\}$, and the vectors

$$\mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \left(\frac{1}{\sqrt{14}}, \frac{2}{\sqrt{14}}, \frac{3}{\sqrt{14}} \right) \quad (+2 \text{ points})$$

$$\mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \left(\frac{1}{\sqrt{6}}, \frac{-2}{\sqrt{6}}, \frac{1}{\sqrt{6}} \right) \quad (+2 \text{ points})$$

form the orthonormal basis.

- (b) The orthogonal projection \mathbf{x} onto the subspace of R^3 spanned by $\{\mathbf{q}_1, \mathbf{q}_2\}$ is

$$\begin{aligned} \text{proj}_{\text{span}\{\mathbf{q}_1, \mathbf{q}_2\}}(\mathbf{x}) &= (\mathbf{x} \cdot \mathbf{q}_1)\mathbf{q}_1 + (\mathbf{x} \cdot \mathbf{q}_2)\mathbf{q}_2 = \frac{-3}{\sqrt{14}} \left(\frac{1}{\sqrt{14}}, \frac{2}{\sqrt{14}}, \frac{3}{\sqrt{14}} \right) + \frac{-3}{\sqrt{6}} \left(\frac{1}{\sqrt{6}}, \frac{-2}{\sqrt{6}}, \frac{1}{\sqrt{6}} \right) \\ &= \frac{-3}{84}(20, -16, 32) = \left(-\frac{5}{7}, \frac{4}{7}, -\frac{8}{7} \right). \quad (+5 \text{ points}) \end{aligned}$$

(Another solution) Let

$$B = \begin{bmatrix} \frac{1}{\sqrt{14}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{14}} & \frac{-2}{\sqrt{6}} \\ \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{6}} \end{bmatrix}.$$

Then, $\text{proj}_{\text{span}\{\mathbf{q}_1, \mathbf{q}_2\}}(\mathbf{x}) = B(B^T B)^{-1} B^T \mathbf{x}$. **(+3 points)** Since $B^T B = I$,

$$\text{proj}_{\text{span}\{\mathbf{q}_1, \mathbf{q}_2\}}(\mathbf{x}) = BB^T \mathbf{x} = \frac{1}{84} \begin{bmatrix} 20 & -16 & 32 \\ -16 & 80 & 8 \\ 32 & 8 & 68 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} -\frac{5}{7} \\ \frac{4}{7} \\ -\frac{8}{7} \end{bmatrix}. \quad (+2 \text{ points})$$

- (c) In (a), we have

$$\mathbf{w}_1 = (\mathbf{w}_1 \cdot \mathbf{q}_1)\mathbf{q}_1 = \sqrt{14}\mathbf{q}_1$$

$$\mathbf{w}_2 = (\mathbf{w}_2 \cdot \mathbf{q}_1)\mathbf{q}_1 + (\mathbf{w}_2 \cdot \mathbf{q}_2)\mathbf{q}_2 = 2\sqrt{14}\mathbf{q}_1 + \sqrt{6}\mathbf{q}_2$$

and this can be written as the QR decomposition of A :

$$A = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{14}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{14}} & \frac{-2}{\sqrt{6}} \\ \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{6}} \end{bmatrix}}_{Q \text{ (+2 points)}} \underbrace{\begin{bmatrix} \sqrt{14} & 2\sqrt{14} \\ 0 & \sqrt{6} \end{bmatrix}}_{R \text{ (+3 points)}}.$$

- (d) The normal equation associated with $Ax = b$ reduces to $Rx = Q^T b$:

$$\begin{bmatrix} \sqrt{14} & 2\sqrt{14} \\ 0 & \sqrt{6} \end{bmatrix} x = \begin{bmatrix} \frac{1}{\sqrt{14}} & \frac{2}{\sqrt{14}} & \frac{3}{\sqrt{14}} \\ \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{14}} \\ \frac{4}{\sqrt{6}} \end{bmatrix}. \text{ (+6 points)}$$

Using the back substitution, we have

$$x_2 = \frac{2}{3}, \quad x_1 = \frac{2}{14} - 2x_2 = -\frac{25}{21}.$$

so a least squares solution is $x = (-\frac{25}{21}, \frac{2}{3})$. **(+4 points)**

□

- Minor calculation mistakes **(-1 points)** or **(-2 points)**

-
- 2** 10+5 (a) Let $A = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$. Find a closest rank one approximation A_1 to A in the Frobenius norm, and report $\|A - A_1\|_F$.
- (b) Let a 3×3 matrix A have orthogonal columns A_1, A_2, A_3 of lengths 2, 3, 4, respectively. Find a singular value decomposition of A .

Solution.

- (a) Since $A^T A = I$, singular values of A are $\sigma_1 = \sigma_2 = 1$. (**+1 points**) Any two vectors that form an orthonormal basis of R^2 can be the right singular vectors v_1 and v_2 of A . Let $v_1 = (1, 0)$ and $v_2 = (0, 1)$. (**+1 points**) Then, the corresponding left singular vectors are

$$\begin{aligned} u_1 &= \frac{1}{\sigma_1} A v_1 = (\cos \theta, \sin \theta) \\ u_2 &= \frac{1}{\sigma_2} A v_2 = (-\sin \theta, \cos \theta) \quad (\text{+3 points}). \end{aligned}$$

Since $\sigma_1 = \sigma_2 = 1$, the following is one of the closest rank one approximations of A :

$$A_1 = \sigma_1 u_1 v_1^T = \begin{bmatrix} \cos \theta & 0 \\ \sin \theta & 0 \end{bmatrix}, \quad (\text{+3 points})$$

and $\|A - A_1\|_F = 1$. (**+2 points**)

- (b) Since $A^T A = \text{diag}(4, 9, 16)$, we have that $\Sigma = \text{diag}(4, 3, 2)$. (**+1 points**) The choice $V = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ (**+2 points**) and $U = \begin{bmatrix} \frac{1}{4}A_3 & \frac{1}{3}A_2 & \frac{1}{2}A_1 \end{bmatrix}$ (**+2 points**) yields a singular value decomposition $A = U\Sigma V^T$.

□

3 Let $A = \begin{bmatrix} 4 & 8 \\ -2 & -4 \\ 1 & 2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$.

(a) Find all the least squares solutions of $A\mathbf{x} = \mathbf{b}$.

(b) Find the minimum norm least squares solution of $A\mathbf{x} = \mathbf{b}$.

Solution.

(a) Since $A^T A = \begin{bmatrix} 21 & 42 \\ 42 & 84 \end{bmatrix}$ and $A^T \mathbf{b} = \begin{bmatrix} 7 \\ 14 \end{bmatrix}$, the solution set of the normal equation $A^T A \mathbf{x} = A^T \mathbf{b}$ (+6 points) is $\mathbf{x} = \left(\frac{1}{3} - 2t, t\right)$. (+4 points)

(Another solution) We want to find \mathbf{x} which minimizes $\|A\mathbf{x} - \mathbf{b}\|$. Since $A\mathbf{x} = \begin{bmatrix} 4 & 8 \\ -2 & -4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} =$

$(x_1 + 2x_2) \begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix}$, it is enough to find k which minimizes $(4k - 1)^2 + (-2k)^2 + (k - 3)^2$.

Then, $k = \frac{1}{3}$ (+6 points) and $\mathbf{x} = \left(\frac{1}{3} - 2t, t\right)$. (+4 points)

(b) A scalar t that minimizes $\|\mathbf{x}(t)\|_2 = \sqrt{\frac{1}{9} - \frac{4}{3}t + 5t^2}$ is $t = \frac{2}{15}$. (+3 points) So, the minimum norm least squares solution is $\mathbf{x} = \left(\frac{1}{15}, \frac{2}{15}\right)$. (+2 points)

□

- Minor calculation mistakes (**-1 points**) or (**-2 points**)
- For (a), writing $\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b}$ instead of $A^T A \mathbf{x} = A^T \mathbf{b}$ (**-2 points**)

4 Let $A = \begin{bmatrix} 3 & 0 & 1 \\ 0 & 3 & 0 \\ 1 & 0 & 3 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ t \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 2t \\ 3 \\ 0 \end{bmatrix}$, and $M(t) = A - \mathbf{u}(t)\mathbf{v}^T(t)$. Note that A has eigenvalues $\lambda_1 = 4$, $\lambda_2 = 3$, $\lambda_3 = 2$.

- (a) Let $\lambda_1(t)$ be the eigenvalue of $M(t)$, where $\lambda_1(0) = 4$. Find the derivative of $\lambda_1(t)$ at $t = 0$.
- (b) Compute $\frac{d}{dt} \log \det M(t)$ at $t = 0$.
- (c) Derive $M^{-1}(t)$ in a form $M^{-1}(t) = A^{-1} + \hat{\mathbf{u}}(t)\hat{\mathbf{v}}^T(t)$ for some vectors $\hat{\mathbf{u}}(t)$ and $\hat{\mathbf{v}}(t)$.
- (d) Let $\hat{\lambda}(t)$ be the eigenvalue of $M^{-1}(t)$, where $\hat{\lambda}(0)$ corresponds to the largest eigenvalue of $M^{-1}(0)$. Find the derivative of $\hat{\lambda}(t)$ at $t = 0$.

Solution.

(a) **Solution 1.**

Note that

$$A - 4I = \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}, A - 3I = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, A - 2I = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Hence the unit eigenvectors of $M(0) = A$ are $\mathbf{x}_1(0) = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$, $\mathbf{x}_2(0) = (0, 1, 0)$, $\mathbf{x}_3(0) = \left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}\right)$ which associate with λ_1 , λ_2 and λ_3 respectively. Since A is a real symmetric matrix, eigenvectors associated with each eigenvalues are linearly independent.

Then

$$X = \begin{bmatrix} \mathbf{x}_1(0) & \mathbf{x}_2(0) & \mathbf{x}_3(0) \end{bmatrix}$$

is an orthogonal matrix.

So, $\mathbf{y}_1(0) = \mathbf{x}_1(0)$ is also the eigenvector of A^T associated with $\lambda_1(0)$, which satisfies $\mathbf{y}_1^T(0)\mathbf{x}_1(0) = 1$. (+3 points) for $\mathbf{x}_1(0), \mathbf{y}_1(0)$)

Since

$$M = \begin{bmatrix} 3 - 2t & -3 & 1 \\ 0 & 3 & 0 \\ 1 - 2t^2 & -3t & 3 \end{bmatrix}$$

we have $\frac{dM}{dt} = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \\ -4t & -3 & 0 \end{bmatrix}$ we have

$$\frac{d\lambda_1}{dt}\Big|_{t=0} = \mathbf{y}_1^T(0) \frac{dM}{dt}\Big|_{t=0} \mathbf{x}_1(0) = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -3 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} = -1.$$

((+5 points) for the formula and (+2 points) for the computation.)

Solution 2.

Considering that

$$M - \lambda I = \begin{bmatrix} 3 - 2t - \lambda & -3 & 1 \\ 0 & 3 - \lambda & 0 \\ 1 - 2t^2 & -3t & 3 - \lambda \end{bmatrix}, \quad (1)$$

we have

$$\det(M - \lambda I) = (3 - \lambda)((3 - 2t - \lambda)(3 - \lambda) - (1 - 2t^2)),$$

$$\lambda = 3 \text{ or } 3 - t \pm \sqrt{1 - t^2}. \quad (+5 \text{ points})$$

If $-1 < t < \frac{1}{\sqrt{2}}$, $3 - t + \sqrt{1 - t^2} \geq 3$ and $3 - t + \sqrt{1 - t^2} \geq 3 - t - \sqrt{1 - t^2}$. Hence we have $\lambda_1(t) = 3 - t + \sqrt{1 - t^2}$ where $-1 < t < \frac{1}{\sqrt{2}}$.

Taking the derivative gives

$$\frac{d(3 - t + \sqrt{1 - t^2})}{dt} = -1 - \frac{t}{\sqrt{1 - t^2}}, \quad (+3 \text{ points}) \quad (2)$$

hence we have $\left. \frac{d\lambda_1}{dt} \right|_{t=0} = -1 \quad (+2 \text{ points})$

(b) **Solution 1.**

Since $\log \det M(t) = \log \lambda_1(t) + \log \lambda_2(t) + \log \lambda_3(t)$, we have

$$\frac{d}{dt} \log \det M(t) = \frac{d}{dt} (\log \lambda_1(t) + \log \lambda_2(t) + \log \lambda_3(t)) = \frac{1}{\lambda_1(t)} \frac{d\lambda_1(t)}{dt} + \frac{1}{\lambda_2(t)} \frac{d\lambda_2(t)}{dt} + \frac{1}{\lambda_3(t)} \frac{d\lambda_3(t)}{dt}.$$

((+3 points) for the formula)

Using that $(0, 1, 0)$ and $\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$ are eigenvectors of A (and A^T) associated with $\lambda_2(0)$ and $\lambda_3(0)$, respectively, we have $\left. \frac{d\lambda_2}{dt} \right|_{t=0} = 0$ and $\left. \frac{d\lambda_3}{dt} \right|_{t=0} = -1$. So,

$$\left. \frac{d}{dt} \log \det M(t) \right|_{t=0} = -\frac{1}{4} - \frac{1}{2} = -\frac{3}{4}.$$

(+2 points)

Solution 2.

$$\det M(t) = 3(8 - 6t + 2t^2). \quad (+2 \text{ points}) \quad (3)$$

Since $8 - 6t + 2t^2 > 0$ for all $t \in \mathbb{R}$, $\log \det M(t)$ is well defined on the real line.

Thus we have

$$\frac{d}{dt} \log \det M(t) = \frac{d}{dt} \log(3(8 - 6t + 2t^2)) = \frac{-6 + 4t}{(8 - 6t + 2t^2)} \quad (+2 \text{ points})$$

so that the answer is $-\frac{3}{4}$. (+1 points)

(c) Note that $A^{-1} = \begin{bmatrix} \frac{3}{8} & 0 & -\frac{1}{8} \\ 0 & \frac{1}{3} & 0 \\ -\frac{1}{8} & 0 & \frac{3}{8} \end{bmatrix}$. By the Sherman-Morrison-Woodbury formula, we have

$$M^{-1}(t) = (A - \mathbf{u}(t)\mathbf{v}^T(t))^{-1} = A^{-1} + A^{-1}\mathbf{u}(t)(1 - \mathbf{v}(t)^T A^{-1}\mathbf{u}(t))^{-1}\mathbf{v}^T(t)A^{-1}.$$

(+3 points)

Since $A^{-1}\mathbf{u}(t) = \left(\frac{3-t}{8}, 0, \frac{3t-1}{8}\right)$ and $1 - \mathbf{v}(t)^T A^{-1}\mathbf{u}(t) = 1 - \frac{3t-t^2}{4}$, we have

$$\begin{aligned} M^{-1}(t) &= \begin{bmatrix} \frac{3}{8} & 0 & -\frac{1}{8} \\ 0 & \frac{1}{3} & 0 \\ -\frac{1}{8} & 0 & \frac{3}{8} \end{bmatrix} + \frac{4}{4-3t+t^2} \begin{bmatrix} \frac{3-t}{8} \\ 0 \\ \frac{3t-1}{8} \end{bmatrix} \begin{bmatrix} \frac{3t}{4} & 1 & -\frac{t}{4} \end{bmatrix} \quad \text{(+2 points)} \\ &= \begin{bmatrix} \frac{3}{8} & 0 & -\frac{1}{8} \\ 0 & \frac{1}{3} & 0 \\ -\frac{1}{8} & 0 & \frac{3}{8} \end{bmatrix} + \begin{bmatrix} \frac{3-t}{2(4-3t+t^2)} \\ 0 \\ \frac{3t-1}{2(4-3t+t^2)} \end{bmatrix} \begin{bmatrix} \frac{3t}{4} & 1 & -\frac{t}{4} \end{bmatrix} \text{ or} \\ &= \begin{bmatrix} \frac{3}{8} & 0 & -\frac{1}{8} \\ 0 & \frac{1}{3} & 0 \\ -\frac{1}{8} & 0 & \frac{3}{8} \end{bmatrix} + \begin{bmatrix} \frac{3-t}{8} \\ 0 \\ \frac{3t-1}{8} \end{bmatrix} \begin{bmatrix} \frac{3t}{4-3t+t^2} & \frac{4}{4-3t+t^2} & -\frac{t}{4-3t+t^2} \end{bmatrix}. \end{aligned}$$

(d) **Solution 1.**

The largest eigenvalue of $M^{-1}(0) = A^{-1}$ is $\hat{\lambda}(0) = \frac{1}{2}$, and its corresponding eigenvector is $\mathbf{x}(0) = \left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}\right)$. $\mathbf{y}(0) = \mathbf{x}(0)$ is also the eigenvalue of $(A^{-1})^T$ associated with the eigenvalue $\hat{\lambda}(0)$, which satisfies $\mathbf{y}^T(0)\mathbf{x}(0) = 1$. **(+3 points)** for $\hat{\lambda}(0), \mathbf{x}, \mathbf{y}$

Using $\frac{dM^{-1}}{dt} = -M^{-1}(t)\frac{dM}{dt}M^{-1}(t)$ **(+5 points)**, we have

$$\begin{aligned} \frac{d\hat{\lambda}}{dt} \Big|_{t=0} &= \mathbf{y}^T(0) \frac{dM^{-1}}{dt} \Big|_{t=0} \mathbf{x}^T(0) = -\mathbf{y}^T(0)A^{-1}\frac{dM}{dt} \Big|_{t=0} A^{-1}\mathbf{x}^T(0) \\ &= - \left[\begin{array}{ccc} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{array} \right] \begin{bmatrix} \frac{3}{8} & 0 & -\frac{1}{8} \\ 0 & \frac{1}{3} & 0 \\ -\frac{1}{8} & 0 & \frac{3}{8} \end{bmatrix} \begin{bmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -3 & 0 \end{bmatrix} \begin{bmatrix} \frac{3}{8} & 0 & -\frac{1}{8} \\ 0 & \frac{1}{3} & 0 \\ -\frac{1}{8} & 0 & \frac{3}{8} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \\ &= - \left[\begin{array}{ccc} \frac{1}{2\sqrt{2}} & 0 & -\frac{1}{2\sqrt{2}} \end{array} \right] \begin{bmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -3 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2\sqrt{2}} \\ 0 \\ -\frac{1}{2\sqrt{2}} \end{bmatrix} = \frac{1}{4} \end{aligned}$$

(+2 points)

Solution 2. From problem (a), the eigenvalues of M are $3, 3-t+\sqrt{1-t^2}$ and $3-t-\sqrt{1-t^2}$. Thus, the largest eigenvalue when t is close to 0 is $\hat{\lambda}(t)$ of M^{-1} is $\frac{1}{3-t-\sqrt{1-t^2}}$. Therefore, by calculating the derivative of $\hat{\lambda}(t)$ gives $\frac{1}{4}$. **(+10 points)**

□

5 Let $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} -4 \\ 10 \\ -20 \end{bmatrix}$.

- 10+5+5+5+5**
- (a) Find a singular value decomposition of A .
 - (b) Find a closest rank one approximation A_1 to A in the Frobenius norm, and compute $\|A - A_1\|_F$. Find a closest rank two approximation A_2 to A , and compute $\|A - A_2\|_F$.
 - (c) Find the pseudoinverse of A .
 - (d) Find a least squares solution of $A\mathbf{x} = \mathbf{b}$ using the pseudoinverse of A in (c).
 - (e) Compute $\|A\|_2$ and $\|A\|_F$.

Solution. (a) Note that $A^T A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$.

Then the eigenvalues of $A^T A$ are $\lambda_1 = 4$ and $\lambda_2 = 2$ with corresponding unit eigenvectors $\mathbf{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$, respectively. (+3 points)

The singular values of A are $\sigma_1 = 2$ and $\sigma_2 = \sqrt{2}$. (+2 points)

Considering that

$$A\mathbf{v}_i = U\Sigma V^T \mathbf{v}_i = \sigma_i \mathbf{u}_i = U\Sigma \mathbf{e}_i, \quad (\text{where } \mathbf{e}_i \text{ is } i\text{-th standard basis of } \mathbf{R}^2)$$

we have $\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ and $\mathbf{u}_2 = \frac{1}{\sigma_2} A\mathbf{v}_2 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$. We choose $\mathbf{u}_3 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$, so that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ form an orthonormal basis for \mathbf{R}^3 . (+3 points)

Then, we have the following singular value decomposition of A :

$$A = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & -1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix}}_U \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{V^T}$$

(+2 points)

(We regarded the reduced SVD as an answer too.)

- (b) A closest rank one approximation to A is

$$A_1 = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T = \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}, \quad (\text{+3 points})$$

and $\|A - A_1\|_F = \sigma_2 = \sqrt{2}$. (+1 points)

Since A has rank 2, A itself is its closest rank two approximation A_2 , so $\|A - A_2\|_F = 0$. (+1 points)

(c) The pseudoinverse of A is

$$A^+ = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_V \underbrace{\begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}}_{\Sigma^+} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & -1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix}}_{U^T} \quad (+3 \text{ points})$$

$$= \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} \end{bmatrix} \quad (+2 \text{ points})$$

(d) Using the pseudoinverse of A , we have a least squares solution:

$$\hat{x} = A^+ b \quad (+4 \text{ points})$$

$$= \begin{bmatrix} -1 \\ -11 \end{bmatrix} \quad (+1 \text{ points}).$$

(e) $\|A\|_2 = \sigma_1 = 2$ (+2 points) and $\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2} = \sqrt{6}$ (+3 points).

□

- The answer without justifications gets no points. (-2 points)

6 Let $A = (a_{ij})$ and $B = (b_{ij})$ be $n \times n$ matrices. State whether or not the following
 $\underline{10+10+5}$ mapping is a matrix norm and verify your answer in (a) and (b).

(a) $A \mapsto \max_{i,j=1,\dots,n} |a_{ij}|$

(b) $A \mapsto \rho(A) = \max_{i=1,\dots,n} |\lambda_i|$, where λ_i 's are eigenvalues of A .

(c) Show that $\rho(A) \leq \|A\| = \max_{\|\mathbf{x}\| \neq 0} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}$.

Solution 1.

(a) It is a matrix norm **(+1 points)**, because it satisfies the following three conditions

- for any nonzero A , $\max_{i,j=1,\dots,n} |a_{ij}| > 0$, **(+3 points)**
- for any scalar c and A , $\max_{i,j=1,\dots,n} |ca_{ij}| = |c| \max_{i,j=1,\dots,n} |a_{ij}|$, **(+3 points)**
- for any matrices A and B ,

$$\max_{i,j=1,\dots,n} |a_{ij} + b_{ij}| \leq \max_{i,j=1,\dots,n} (|a_{ij}| + |b_{ij}|) \leq \max_{i,j=1,\dots,n} |a_{ij}| + \max_{i,j=1,\dots,n} |b_{ij}| \quad \text{(+3 points).}$$

(b) Consider matrices $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$. Then, we have

$$\rho(A + B) = 1 > \rho(A) + \rho(B) = 0 \quad \text{(+9 points),}$$

which does not satisfy the property of a matrix norm. So, $\rho(A)$ is not a matrix norm.
(+1 points)

(c) Let λ_i 's be eigenvalues of A , and \mathbf{v}_i 's be their corresponding eigenvectors. Since $\|A\mathbf{v}_i\| = \|\lambda_i \mathbf{v}_i\| = |\lambda_i| \|\mathbf{v}_i\|$, **(+2 points)** we have

$$|\lambda_i| = \frac{\|A\mathbf{v}_i\|}{\|\mathbf{v}_i\|} \leq \|A\| \quad \text{(+3 points),}$$

and this concludes the proof. □

- The answer without justifications gets no points.

7 Consider a set of m data points $\mathbf{z}_i = (x_i, y_i)$ for $i = 1, \dots, m$, where $\sum_{i=1}^m x_i = 10+5$ and $\sum_{i=1}^m y_i = 0$.

- (a) Describe a way to find a line $y(x) = ax$ such that the sum of squared distances from data points to the line is minimized. Justify your answer.
- (b) Find the line $y(x) = ax$ in (a) for data points $\{(-1, -1), (0, 1), (1, 0)\}$, using the approach in (a).

Solution 1.

- (a) Let $\{\mathbf{u}_1, \mathbf{u}_2\}$ be an orthonormal basis for \mathbb{R}^2 , where $\mathbf{u}_1 = \left(\frac{1}{\sqrt{a^2+1}}, \frac{a}{\sqrt{a^2+1}} \right)$ denotes the direction of a line $y(x)$. (+2 points) We want to find \mathbf{u}_1 and \mathbf{u}_2 that minimize $\sum_{i=1}^m |\mathbf{z}_i \mathbf{u}_2|^2$, which is the sum of squared distances from data points to the line in the direction of \mathbf{u}_1 . (+2 points) Let $A = [\mathbf{z}_1 \ \cdots \ \mathbf{z}_m]$, and we have

$$\|A\|_F^2 = \sum_{i=1}^m \|\mathbf{z}_i\|^2 = \sum_{i=1}^m |\mathbf{z}_i^T \mathbf{u}_1|^2 + \sum_{i=1}^m |\mathbf{z}_i^T \mathbf{u}_2|^2 = \mathbf{u}_1^T A A^T \mathbf{u}_1 + \mathbf{u}_2^T A A^T \mathbf{u}_2 \quad (+3 \text{ points}).$$

Since $\|A\|_F^2$ is fixed, minimizing $\sum_{i=1}^m |\mathbf{z}_i \mathbf{u}_2|^2 = \mathbf{u}_2^T A A^T \mathbf{u}_2$ is equivalent to maximizing $\mathbf{u}_1^T A A^T \mathbf{u}_1$, which corresponds to finding a unit eigenvector of $A A^T$ corresponding to its largest eigenvalue. (+3 points)

- (b) The eigenvalues of $A A^T = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ are 3 and 1. The unit eigenvector of $A A^T$ associated with the largest eigenvalue 3 is $\mathbf{u}_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right)$. (+3 points) So, the line in (a) for the given data points is $y(x) = x$. (+2 points)

□

- The answer without justifications gets no points.
- You have -1 point for each calculation mistakes

- 8** Consider minimizing a quadratic function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{b}^T \mathbf{x}$ for an $n \times n$ symmetric matrix $A = [A_1 \ \dots \ A_n]$, which is equivalent to finding an $n \times 1$ vector solution of $A\mathbf{x} = \mathbf{b}$. A gradient descent method with $\nabla f(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$ finds the solution by iteratively updating $\mathbf{x}^{(i)} = (x_1^{(i)}, \dots, x_n^{(i)})$ as

$$\mathbf{x}^{(i+1)} = g(\mathbf{x}^{(i)}) = \mathbf{x}^{(i)} - \alpha \nabla f(\mathbf{x}^{(i)}) = \mathbf{x}^{(i)} - \alpha(A\mathbf{x}^{(i)} - \mathbf{b})$$

for $i = 0, 1, \dots$, with appropriately chosen α . For large n , computing $A\mathbf{x}^{(i)}$ is expensive, so we want to approximate $A\mathbf{x}^{(i)}$ by $C\mathbf{r}$, a product of an $n \times s$ matrix C and an $s \times 1$ vector \mathbf{r} for $s \ll n$. Here, the columns of C consists of randomly chosen s columns of A (with replacement) and \mathbf{r} consists of corresponding s elements of $\mathbf{x}^{(i)}$. Let p_k denote the probability of choosing the k th column of A . Then, an approximation of gradient descent update becomes

$$\mathbf{x}^{(i+1)} = \tilde{g}(\mathbf{x}^{(i)}) = \mathbf{x}^{(i)} - \alpha(C\mathbf{r} - \mathbf{b}).$$

- (a) Modify $C\mathbf{r}$ (by multiplying appropriate scalar) so that $E[C\mathbf{r}] = A\mathbf{x}^{(i)}$. Let $k(j)$ denote the index randomly chosen at the j th trial. Verify your answer.
- (b) State the probability distribution p_1, \dots, p_n that minimizes $E[\|g(\mathbf{x}^{(i)}) - \tilde{g}(\mathbf{x}^{(i)})\|_2^2]$ for $C\mathbf{r}$ in (a).

- (c) Let $A = \begin{bmatrix} \sqrt{6} & 3 & -1 \\ 3 & 4 & 0 \\ -1 & 0 & \sqrt{3} \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$ and $\mathbf{x}^{(i)} = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}$. Consider the probability distribution $p_k = \frac{\|A_k\|_2}{\sum_{k=1}^n \|A_k\|_2}$ for all k . Compute the values of p_k and $E[\|g(\mathbf{x}^{(i)}) - \tilde{g}(\mathbf{x}^{(i)})\|_2^2]$ for $\alpha = 1$ and $s = 2$.

Solution 1.

- (a) Note that $C\mathbf{r} = \sum_{j=1}^s A_{k(j)} x_{k(j)}^{(i)}$. At the j th trial, multiplying $A_{k(j)} x_{k(j)}^{(i)}$ by $\frac{1}{sp_{k(j)}}$ yields

$$C\mathbf{r} = \sum_{j=1}^s \frac{A_{k(j)} x_{k(j)}^{(i)}}{sp_{k(j)}} \quad (+4 \text{ points}),$$

and this satisfies

$$E[C\mathbf{r}] = \sum_{j=1}^s \sum_{k=1}^n p_k \frac{A_k x_k^{(i)}}{sp_k} = A\mathbf{x}^{(i)} \quad (+1 \text{ points}).$$

- (b) We have that

$$E[\|g(\mathbf{x}^{(i)}) - \tilde{g}(\mathbf{x}^{(i)})\|_2^2] = \alpha^2 E[\|C\mathbf{r} - A\mathbf{x}^{(i)}\|_2^2] = \alpha^2 \left(\sum_{k=1}^n \frac{\|A_k\|_2^2 \|x_k^{(i)}\|^2}{sp_k} - \frac{1}{s} \|A\mathbf{x}^{(i)}\|_2^2 \right) \quad (+3 \text{ points}).$$

The probability distribution that minimizes the above is

$$p_k = \frac{\|A_k\|_2 \|x_k^{(i)}\|_2}{\sum_{k=1}^n \|A_k\|_2 \|x_k^{(i)}\|_2} \quad (+2 \text{ points}).$$

(c) Since $\|A_1\|_2 = 4$, $\|A_2\|_2 = 5$, and $\|A_3\|_2 = 2$, we have

$$p_1 = \frac{4}{11}, \quad p_2 = \frac{5}{11}, \quad p_3 = \frac{2}{11} \text{ (+2 points)}$$

and for such distribution we have

$$E[\|g(\mathbf{x}^{(i)}) - \tilde{g}(\mathbf{x}^{(i)})\|_2^2] = \frac{16 \times 1}{2 \times \frac{4}{11}} + \frac{25 \times 4}{2 \times \frac{5}{11}} + \frac{4 \times 0}{2 \times \frac{2}{11}} - \frac{1}{2}(68 - 12\sqrt{6}) = 98 + 6\sqrt{6} \text{ (+8 points).}$$

□

- The answer without justifications gets no points.
- You have -1 point for each calculation mistakes.