Problem 1: Orthogonal Projection onto a Subspace

Problem: Let the subspace S of $\mathbb{R}^3$ be spanned by

$$u_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}.$$

Given the vector

$$v = \begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix},$$

a) Find the orthogonal projection v_S of v onto S. That is, determine scalars a and b such that

$$v_S = a\, u_1 + b\, u_2,$$

and the error vector $v_\perp = v - v_S$ is orthogonal to every vector in S.

b) Verify that $v_\perp$ is orthogonal to both u_1 and u_2.

Solution:

a) To determine a and b, we require that

$$(v - (a\, u_1 + b\, u_2)) \cdot u_1 = 0 \quad \text{and} \quad (v - (a\, u_1 + b\, u_2)) \cdot u_2 = 0.$$

First, compute the necessary dot products:

$$u_1 \cdot u_1 = 1^2 + 2^2 + 1^2 = 6, \quad u_2 \cdot u_2 = 0^2 + 1^2 + (-1)^2 = 2,$$
$$u_1 \cdot u_2 = 1 \cdot 0 + 2 \cdot 1 + 1 \cdot (-1) = 1.$$

Also,

$$v \cdot u_1 = 4 \cdot 1 + 3 \cdot 2 + 2 \cdot 1 = 12,$$
$$v \cdot u_2 = 4 \cdot 0 + 3 \cdot 1 + 2 \cdot (-1) = 1.$$

This yields the system:

$$\begin{cases} 6a + b = 12, \\ a + 2b = 1. \end{cases}$$

Solve the second equation for a:

$$a = 1 - 2b.$$

Substitute into the first equation:

$$6(1 - 2b) + b = 12 \quad \Longrightarrow \quad 6 - 12b + b = 12,$$
$$6 - 11b = 12 \quad \Longrightarrow \quad -11b = 6 \quad \Longrightarrow \quad b = -\frac{6}{11}.$$

Then,

$$a = 1 - 2\left(-\frac{6}{11}\right) = 1 + \frac{12}{11} = \frac{23}{11}.$$

Therefore, the projection is:

$$v_S = \frac{23}{11} u_1 - \frac{6}{11} u_2.$$

b) The error vector is

$$v_\perp = v - v_S.$$

By construction, the coefficients a and b ensure that

$$v_\perp \cdot u_1 = 0 \quad \text{and} \quad v_\perp \cdot u_2 = 0.$$

(A direct calculation confirms that both dot products vanish.)

Problem 2: Linear Dependence and Orthogonality in $\mathbb{R}^4$

Problem: Consider the vectors in $\mathbb{R}^4$:

$$a = \begin{bmatrix} 2 \\ -1 \\ 0 \\ 3 \end{bmatrix}, \quad b = \begin{bmatrix} 4 \\ -2 \\ 1 \\ 6 \end{bmatrix}, \quad c = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}.$$

a) Show that the vectors a, b, and c are linearly dependent and express c as a linear combination of a and b.

b) Let

$$d = \begin{bmatrix} k \\ 0 \\ 2 \\ -1 \end{bmatrix}.$$

Determine the value of k such that d is orthogonal to a.

Solution:

a) Notice that

$$2a = \begin{bmatrix} 4 \\ -2 \\ 0 \\ 6 \end{bmatrix}.$$

Then, compute

$$b - 2a = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

Observe that

$$c = -\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = -(b - 2a).$$

Hence, we can express c as:

$$c = 2a - b.$$

This shows that a, b, and c are linearly dependent.

b) For d to be orthogonal to a, we require:

$$d \cdot a = 0.$$

Compute the dot product:

$$d \cdot a = 2k + 0 + 0 - 3 = 2k - 3.$$

Setting this equal to zero gives:

$$2k - 3 = 0 \implies k = \frac{3}{2}.$$

Problem 3: Eigenvalues and Eigenvectors of Matrix B

Consider the matrix

$$B = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}.$$

a) Eigenvalues

To find the eigenvalues, we compute the characteristic polynomial:

$$\det(\lambda I - B) = \det \begin{bmatrix} \lambda - 3 & -1 & -1 \\ -1 & \lambda - 3 & -1 \\ -1 & -1 & \lambda - 3 \end{bmatrix}.$$

It can be shown (by expansion or using symmetry arguments) that the eigenvalues are:

$$\lambda_1 = 5 \quad \text{and} \quad \lambda_2 = \lambda_3 = 2.$$

b) Eigenvectors and Diagonalizability

For $\lambda_1 = 5$: Solving $(B - 5I)x = 0$ leads to

$$\begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix} x = 0.$$

A suitable eigenvector is:

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

For $\lambda_2 = 2$: Solving $(B - 2I)x = 0$ yields

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} x = 0,$$

which implies

$$x_1 + x_2 + x_3 = 0.$$

Two independent eigenvectors satisfying this condition can be chosen as:

$$v_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad v_3 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

Since we have three linearly independent eigenvectors, the matrix B is diagonalizable. In other words, there exists an invertible matrix P (whose columns are v_1, v_2, and v_3) such that

$$D = P^{-1}BP$$

is diagonal, with

$$D = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

c) Rank of B

Since none of the eigenvalues is zero, the matrix B is full rank:

$$\text{rank}(B) = 3.$$

d) Trace and Sum of Eigenvalues

The trace of B is

$$\mathrm{tr}(B) = 3 + 3 + 3 = 9.$$

The sum of the eigenvalues is also

$$5 + 2 + 2 = 9.$$

Thus, the trace equals the sum of the eigenvalues.

e) Determinant and Product of Eigenvalues

The determinant of B is

$$\det(B) = 5 \times 2 \times 2 = 20.$$

This is equal to the product of the eigenvalues:

$$5 \cdot 2 \cdot 2 = 20.$$

Problem 4: Linear Transformations and Image Warping

Geometric transformations are widely used in computer vision and graphics to manipulate images. We explore transformations using homogeneous coordinates and linear transformation matrices. Homogeneous coordinates are a system used in mathematics and computer graphics to represent points in a projective space. They extend traditional Cartesian coordinates by adding an extra dimension, simplifying transformations like translation, rotation, scaling, and perspective projection. As a key idea, in Cartesian Coordinates (2D), a point is represented as (x, y), while in Homogeneous Coordinates (2D), the same point is represented as (x, y, w), where w is a scalar "weight" (usually 1 for finite points).

a) **Shearing Transformation**

A shear transformation distorts a shape by shifting one axis while keeping the other fixed.

- The transformation matrix for a horizontal shear (shifting x in proportion to y) with shear factor m:

$$S_x = \begin{bmatrix} 1 & m & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

This maps (x, y) to $(x + my, y)$.

- The transformation matrix for a vertical shear (shifting y in proportion to x) with shear factor n:

$$S_y = \begin{bmatrix} 1 & 0 & 0 \\ n & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

This maps (x, y) to $(x, y + nx)$.

b) **Perspective Projection**

Perspective transformations introduce depth effects. Suppose a point (x, y) is projected onto a plane such that:

$$x' = \frac{x}{1+z}, \quad y' = \frac{y}{1+z}.$$

In homogeneous coordinates, this transformation can be represented by the matrix:

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1+z \end{bmatrix}.$$

After applying P, the resulting coordinates are normalized by dividing by the third coordinate.

c) **Composite Transformation in Graphics**

Given a sequence of transformations:

(a) Horizontal shear with $m = 0.5$.
(b) Scaling: doubling width, halving height.
(c) Translation by $(2, -1)$.

The corresponding matrices are:

$$S_x = \begin{bmatrix} 1 & 0.5 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad T = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

The final transformation matrix is their product:

$$M = TDS_x.$$

Performing the matrix multiplication:

$$M = \begin{bmatrix} 2 & 1 & 2 \\ 0 & 0.5 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

This single matrix applies all transformations in sequence.

Problem 5: Calculus and the Chain Rule

Compute the derivative $\frac{df}{dx}$ for the following functions using the chain rule. Provide the dimensions of each partial derivative and describe your steps in detail.

a) $f(z) = e^{-z}, \quad z = x^T W x, \quad x \in \mathbb{R}^N, \quad W \in \mathbb{R}^{N \times N}$ is a symmetric matrix.

b) $f = \sigma(z), \quad z = Ax + b, \quad x \in \mathbb{R}^N, \quad A \in \mathbb{R}^{M \times N}, \quad b \in \mathbb{R}^M, \quad f \in \mathbb{R}^M$, where $\sigma(z)$ is the elementwise sigmoid function.

Solution:

a) We use the chain rule:

$$\frac{df}{dx} = \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} \in \mathbb{R}^{1 \times N}.$$

Compute:

$$\frac{\partial f}{\partial z} = -e^{-z} \quad \text{(scalar derivative)},$$

$$\frac{\partial z}{\partial x} = (W + W^T)x \in \mathbb{R}^{1 \times N}.$$

Since W is symmetric, $W + W^T = 2W$. Thus,

$$\frac{df}{dx} = -e^{-x^T W x}(2Wx).$$

b) Again, using the chain rule:

$$\frac{df}{dx} = \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} \in \mathbb{R}^{M \times N}.$$

Compute:

$$\frac{\partial f}{\partial z} = \operatorname{diag}(\sigma(z)(1 - \sigma(z))) \in \mathbb{R}^{M \times M},$$

$$\frac{\partial z}{\partial x} = A \in \mathbb{R}^{M \times N}.$$

Hence,

$$\frac{df}{dx} = \operatorname{diag}(\sigma(z)(1 - \sigma(z)))\, A \in \mathbb{R}^{M \times N}.$$

Problem 6: Unconstrained Quadratic Minimization

Consider the quadratic function

$$f(x) = \frac{1}{2}x^T Qx + c^T x + d,$$

where:

- $Q \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix,
- $c \in \mathbb{R}^n$ is a given vector,
- $d \in \mathbb{R}$ is a constant, and
- $x \in \mathbb{R}^n$ is the optimization variable.

Answer the following questions:

a) **Gradient and Hessian:** Compute the gradient $\nabla f(x)$ and the Hessian $\nabla^2 f(x)$.

b) **Minimizer:** Find the unique minimizer x^* of $f(x)$.

c) **Uniqueness:** Explain why the positive definiteness of Q guarantees that x^* is the unique minimizer.

Solution:

a) The function is given by

$$f(x) = \frac{1}{2}x^T Qx + c^T x + d.$$

The gradient is computed as:

$$\nabla f(x) = Qx + c.$$

Since Q is constant (and symmetric), the Hessian is:

$$\nabla^2 f(x) = Q.$$

b) To find the minimizer, set the gradient equal to zero:

$$Qx + c = 0.$$

Solving for x yields:

$$x^* = -Q^{-1}c.$$

c) Because Q is symmetric positive definite, it is invertible and the quadratic function $f(x)$ is strictly convex. Strict convexity ensures that the stationary point is the unique global minimizer. Thus, $x^* = -Q^{-1}c$ is the unique minimizer of $f(x)$.

Problem 7: Weighted Least Squares Quadratic Curve Fitting

Consider the following data points:

$$(0, 2), \quad (1, 2.5), \quad (2, 3.6), \quad (3, 5.1), \quad (4, 8.2).$$

We wish to fit a quadratic model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

to these data using the weighted least squares method. Suppose that the measurements corresponding to the points with $x = 2$ and $x = 3$ are less reliable; therefore, they are assigned half the weight of the other points. Formulate the weighted least squares problem in matrix form and determine the coefficients β_0, β_1, and β_2.

Solution:

a) **Formulation:**
Define the design matrix X, the parameter vector $\boldsymbol{\beta}$, and the observation vector y as:

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad y = \begin{bmatrix} 2 \\ 2.5 \\ 3.6 \\ 5.1 \\ 8.2 \end{bmatrix}.$$

In weighted least squares, we minimize

$$\|W(X\boldsymbol{\beta} - y)\|_2^2,$$

where W is a diagonal matrix of weights. Assign the weights:

$$w_1 = 1, \quad w_2 = 1, \quad w_3 = 0.5, \quad w_4 = 0.5, \quad w_5 = 1.$$

It is common to work with the square roots of the weights when forming the weighted design matrix. Thus, define:

$$W = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.7071 & 0 & 0 \\ 0 & 0 & 0 & 0.7071 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

b) **Normal Equations:**
The weighted least squares solution $\boldsymbol{\beta}^*$ satisfies

$$(X^T W^2 X)\boldsymbol{\beta} = X^T W^2 y,$$

where $W^2 = \text{diag}(1,\, 1,\, 0.5,\, 0.5,\, 1)$.

For the 5 data points with

$$x_1 = 0, \quad x_2 = 1, \quad x_3 = 2, \quad x_4 = 3, \quad x_5 = 4,$$

compute:

$$\sum_{i=1}^{5} w_i = 4, \quad \sum_{i=1}^{5} w_i x_i = 7.5, \quad \sum_{i=1}^{5} w_i x_i^2 = 23.5,$$

$$\sum_{i=1}^{5} w_i x_i^3 = 82.5, \quad \sum_{i=1}^{5} w_i x_i^4 = 305.5.$$

Thus,

$$X^T W^2 X = \begin{bmatrix} 4 & 7.5 & 23.5 \\ 7.5 & 23.5 & 82.5 \\ 23.5 & 82.5 & 305.5 \end{bmatrix}.$$

Next, compute:

$$X^T W^2 y = \begin{bmatrix} 17.05 \\ 46.55 \\ 164.85 \end{bmatrix}.$$

The normal equations are:

$$\begin{bmatrix} 4 & 7.5 & 23.5 \\ 7.5 & 23.5 & 82.5 \\ 23.5 & 82.5 & 305.5 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 17.05 \\ 46.55 \\ 164.85 \end{bmatrix}.$$

c) **Solution for the Coefficients:**
Solving the above system yields approximately:

$$\beta_0 \approx 2.23, \quad \beta_1 \approx -0.455, \quad \beta_2 \approx 0.491.$$

Therefore, the fitted quadratic curve is:

$$y \approx 2.23 - 0.455x + 0.491x^2.$$

d) **Interpretation:**
By assigning half the weight to the points with $x = 2$ and $x = 3$, their influence is reduced, yielding a quadratic curve that better reflects the more reliable data.

Problem 8: Optimal Allocation of Advertising Budget

A company has an advertising budget of \$400,000 that can be allocated between two advertising channels: TV and online. Let

$$x \quad \text{(in thousands of dollars)}$$

be the amount spent on TV advertising, and

$$y \quad \text{(in thousands of dollars)}$$

be the amount spent on online advertising. The expected sales (in thousands of units) are modeled by

$$S(x, y) = 50\sqrt{x} + 80\sqrt{y}.$$

Determine the optimal allocation (x^*, y^*) that maximizes expected sales subject to the budget constraint

$$x + y = 400.$$

Solve the problem using the method of Lagrange multipliers.

Solution:

a) **Lagrangian Formulation:**
Define

$$\mathcal{L}(x, y, \lambda) = 50\sqrt{x} + 80\sqrt{y} + \lambda(400 - x - y).$$

b) **First-Order Conditions:**
Compute:

- With respect to x:

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{50}{2\sqrt{x}} - \lambda = 0 \quad \Longrightarrow \quad \lambda = \frac{25}{\sqrt{x}}.$$

- With respect to y:

$$\frac{\partial \mathcal{L}}{\partial y} = \frac{80}{2\sqrt{y}} - \lambda = 0 \quad \Longrightarrow \quad \lambda = \frac{40}{\sqrt{y}}.$$

- With respect to λ:

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 400 - x - y = 0 \quad \Longrightarrow \quad x + y = 400.$$

c) **Solve for x and y:**
Equate:

$$\frac{25}{\sqrt{x}} = \frac{40}{\sqrt{y}} \quad \Longrightarrow \quad 25\sqrt{y} = 40\sqrt{x}.$$

Dividing by 5:

$$5\sqrt{y} = 8\sqrt{x} \quad \Longrightarrow \quad \sqrt{y} = \frac{8}{5}\sqrt{x}.$$

Squaring gives:

$$y = \frac{64}{25}x.$$

Substitute into $x + y = 400$:

$$x + \frac{64}{25}x = 400 \quad \Longrightarrow \quad x\left(1 + \frac{64}{25}\right) = 400.$$

Note that:

$$1 + \frac{64}{25} = \frac{89}{25}.$$

Hence,

$$x = \frac{400 \cdot 25}{89} \approx 112.36,$$

and

$$y = 400 - 112.36 \approx 287.64.$$

d) **Interpretation:**
The optimal allocation is approximately:

$$x^* \approx 112.36 \quad (\$112,360 \text{ on TV advertising}),$$
$$y^* \approx 287.64 \quad (\$287,640 \text{ on online advertising}).$$

The estimated maximum expected sales are:

$$S(x^*, y^*) \approx 50\sqrt{112.36} + 80\sqrt{287.64} \approx 1888 \quad \text{(thousands of units)}.$$

Problem 9: Sums and Conditionals for Poisson Random Variables

Consider two independent Poisson random variables X and Y with parameters $\lambda = 2$ and $\mu = 3$, respectively. Define

$$Z = X + Y.$$

Answer the following:

a Find the probability mass function (pmf) of Z.

b Compute $P(X = 1 \mid Z = 4)$.

Solution:

a **Distribution of Z:**
Since X and Y are independent Poisson random variables, Z is Poisson with parameter

$$\lambda + \mu = 5.$$

Thus,

$$P(Z = k) = e^{-5} \frac{5^k}{k!}, \quad k = 0, 1, 2, \dots.$$

b **Conditional Probability:**
For independent Poisson random variables,

$$X \mid (Z = n) \sim \text{Binomial}\left(n, \frac{\lambda}{\lambda + \mu}\right).$$

With $\lambda = 2$ and $\mu = 3$,

$$X \mid (Z = 4) \sim \text{Binomial}\left(4, \frac{2}{5}\right).$$

Hence,

$$P(X = 1 \mid Z = 4) = \binom{4}{1} \left(\frac{2}{5}\right) \left(\frac{3}{5}\right)^3 = 4 \cdot \frac{2}{5} \cdot \frac{27}{125} = \frac{216}{625} \approx 0.3456.$$

Problem 10: Low-Rank Approximation and Pseudoinverse via SVD

Consider the matrix

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

Answer the following:

a **SVD Decomposition:** Compute the singular value decomposition of A such that

$$A = USV^T,$$

where U is a 2×2 orthogonal matrix, S is a 2×3 diagonal matrix (with non-negative singular values on the diagonal), and V is a 3×3 orthogonal matrix.

b **Rank-1 Approximation:** Use the SVD of A to form the best rank-1 approximation A_1 (in the Frobenius norm) by retaining only the largest singular value.

c **Pseudoinverse:** Compute the Moore-Penrose pseudoinverse A^+ of A using its SVD.

d **Least-Squares Solution:** Solve the least-squares problem $Ax = b$ for

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

by computing $x = A^+ b$.

Solution:

a The SVD of A is given by

$$A = USV^T,$$

where:

- U is a 2×2 orthogonal matrix,
- S is a 2×3 diagonal matrix with singular values σ_1 and σ_2 (with $\sigma_1 \geq \sigma_2 \geq 0$) on the diagonal, and
- V is a 3×3 orthogonal matrix.

Using a numerical tool (e.g., `np.linalg.svd` in Python), one obtains approximate values:

$$U \approx \begin{bmatrix} -0.4046 & -0.9145 \\ -0.9145 & 0.4046 \end{bmatrix}, \quad S \approx \begin{bmatrix} 9.5080 & 0 & 0 \\ 0 & 0.7729 & 0 \end{bmatrix}, \quad V^T \approx \begin{bmatrix} -0.4287 & -0.5663 & -0.7039 \\ 0.8059 & 0.1124 & -0.5810 \\ 0.4082 & -0.8165 & 0.4082 \end{bmatrix}.$$

b The best rank-1 approximation A_1 is obtained by retaining only the largest singular value σ_1 along with its corresponding singular vectors u_1 (first column of U) and v_1 (first row of V^T). That is,

$$A_1 = \sigma_1 \, u_1 v_1^T.$$

Numerically,

$$A_1 \approx 9.5080 \begin{bmatrix} -0.4046 \\ -0.9145 \end{bmatrix} \begin{bmatrix} -0.4287 & -0.5663 & -0.7039 \end{bmatrix}.$$

c The Moore-Penrose pseudoinverse A^+ is computed as

$$A^+ = VS^+U^T,$$

where S^+ is obtained by taking the reciprocal of the nonzero singular values:

$$S = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix} \implies S^+ = \begin{bmatrix} 1/\sigma_1 & 0 \\ 0 & 1/\sigma_2 \\ 0 & 0 \end{bmatrix}.$$

Hence,

$$A^+ \approx V \, S^+ \, U^T.$$

d To solve $Ax = b$ for $b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, compute:

$$x = A^{+}b.$$

Using the computed pseudoinverse A^{+} from part (c) gives the least-squares solution.

A sample Python code to perform these computations is provided below:

```python
import numpy as np

# Define the matrix A and vector b
A = np.array([[1, 2, 3],
              [4, 5, 6]])
b = np.array([1, 0])

# Compute the SVD of A
U, s, Vt = np.linalg.svd(A, full_matrices=False)
print("U =", U)
print("Singular values =", s)
print("Vt =", Vt)

# Construct the diagonal matrix S from singular values
S = np.diag(s)

# Form the rank-1 approximation by keeping only the largest singular value
k = 1
A1 = s[0] * np.outer(U[:, 0], Vt[0, :])
print("Rank-1 approximation A1 =\n", A1)

# Compute the pseudoinverse of A
S_inv = np.diag(1/s)
A_pinv = Vt.T @ S_inv @ U.T
print("Pseudoinverse A+ =\n", A_pinv)

# Solve the least-squares problem Ax = b
x = A_pinv @ b
print("Solution x =", x)
```