

Least squares, projections and the SVD

Ten problems on projections, least squares and the singular value decomposition: projecting onto a line and a column space, fitting a line through three points, why the residual is orthogonal, an SVD computed by hand, reading rank and null space off it, the pseudoinverse and the best rank-one approximation, with worked solutions and the mistakes that confuse U with V or drop $(A^T A)^{-1}$.

BEFORE YOU START

A system $Ax = b$ with more equations than unknowns usually has no solution; least squares drops b perpendicularly onto the column space of A and takes the foot of the perpendicular as the best fit. These ten problems build that projection, fit a line with it, then compute an SVD by hand and read the same answers off it.

- Vectors are columns, like $(1, 2, 2)^T$; $\|w\| = \sqrt{w^T w}$, and w, z are orthogonal when $w^T z = 0$.
- $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times k}$ (n rows, k columns, $n \geq k$). $\text{col}(A)$ is the set of all Ax ; $\text{col}(A)^\perp$ is the set of vectors orthogonal to every column of A .
- P is a projection matrix: Pb is the point of a line or a column space closest to b . $\hat{x}$ is the least-squares solution, the x that minimises $\|Ax - b\|$, so $Pb = A\hat{x}$; $r = b - A\hat{x}$ is the residual.
- The matrix-calculus page derived the normal equations $A^T A \hat{x} = A^T b$ from the gradient of $\|Ax - b\|^2$, and showed $A^T A$ is invertible exactly when the columns of A are independent; this page reaches them by geometry, and then by the SVD.
- The singular value decomposition (SVD) is $A = U \Sigma V^T$ with $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_k)$, $\sigma_1 \geq \sigma_2 \geq \dots \geq 0$ the singular values; V is a $k \times k$ orthogonal matrix, $V^T V = V V^T = I$, with columns v_i ; U is $n \times k$ with orthonormal columns u_i , $U^T U = I$. Column by column, $AV = U \Sigma$ says $Av_i = \sigma_i u_i$. (This is the thin SVD; the full one pads U to square and Σ with zero rows.)
- Σ^+ is Σ with each nonzero σ_i replaced by $1/\sigma_i$, and $A^+ = V \Sigma^+ U^T$ is the pseudoinverse. $\|M\|_2$, the spectral norm, is the largest $\|Mw\|$ over unit vectors w ; it equals $\sigma_1(M)$.

Builds on: [Eigenvalues and eigenvectors by hand](#)

Problems

1. Find the projection p of b onto the line through a , and the matrix P with $p = Pb$. difficulty ·

2. Show that P from Problem 1 satisfies $P^2 = P$ and $P^\top = P$, has rank 1, and has eigenvalues 1 and 0. What are the eigenvectors? difficulty ··

3. For A with independent columns, find the matrix that projects onto $\text{col}(A)$, and show the residual $b - Pb$ is orthogonal to every column of A . difficulty ··

4. Fit the line $y = c + mx$ to the points $(0, 1)$, $(1, 2)$, $(2, 2)$ by least squares: set up A and b , form the normal equations and solve them. difficulty ··

5. Compute the residual $r = b - A\hat{x}$ of Problem 4 and verify $A^\top r = 0$. difficulty ··

6. Compute the SVD of $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$ by hand: singular values, V , then U . difficulty ···

7. A 3×2 matrix has singular values $\sigma_1 > 0$ and $\sigma_2 = 0$. Describe its rank, column space and null space in terms of u_1 and v_2 . difficulty ··

8. Define $A^+ = V\Sigma^+U^\top$. Show that when A has independent columns, $A^+ = (A^\top A)^{-1}A^\top$, and compute A^+b for Problem 4. difficulty ···

9. For A of Problem 6, write down the best rank-one approximation A_1 and the size of the error $\|A - A_1\|_2$. difficulty ··

10. Show $I - P$ projects onto the orthogonal complement of $\text{col}(A)$, and that $\|b\|^2 = \|Pb\|^2 + \|(I - P)b\|^2$. difficulty ···

Answers

1. $p = \frac{a^\top b}{a^\top a} a, P = \frac{a a^\top}{a^\top a}$
2. $P^2 = P$ and $P^\top = P$ by direct multiplication; rank 1 (every column is a multiple of a); $Pa = a$ so $\lambda = 1$ with eigenvector a , and $Px = 0$ for every $x \perp a$, so $\lambda = 0$ with multiplicity $n - 1$
3. $P = A(A^\top A)^{-1}A^\top$; $A^\top(b - Pb) = A^\top b - A^\top b = 0$
4. $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, b = (1, 2, 2)^\top$; $A^\top A = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}, A^\top b = (5, 6)^\top$; $c = \frac{7}{6}, m = \frac{1}{2}$
5. $r = (-\frac{1}{6}, \frac{1}{3}, -\frac{1}{6})^\top$ and $A^\top r = (0, 0)^\top$
6. $\sigma_1 = 3\sqrt{5}, \sigma_2 = \sqrt{5}$; $V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$; $u_i = Av_i/\sigma_i$ gives $U = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 3 \\ 3 & -1 \end{bmatrix}$; check $U\Sigma V^\top = A$
7. rank 1; $\text{col}(A) = \text{span}\{u_1\}$; null space = $\text{span}\{v_2\}$, because $Av_2 = \sigma_2 u_2 = 0$
8. $(A^\top A)^{-1}A^\top = V\Sigma^{-1}U^\top = A^+$; $A^+b = (\frac{7}{6}, \frac{1}{2})^\top$
9. $A_1 = \sigma_1 u_1 v_1^\top$ and $\|A - A_1\|_2 = \sigma_2 = \sqrt{5}$
10. $I - P$ is the projection onto $\text{col}(A)^\perp$, and $\|b\|^2 = \|Pb\|^2 + \|(I - P)b\|^2$

Worked solutions

Problem 1

Find the projection p of b onto the line through a , and the matrix P with $p = Pb$.

Here $a, b \in \mathbb{R}^n$ and $a \neq 0$.

$p = ta$ for some number t .	p lies on the line through a , and every point of that line is a multiple of a .
$a^\top(b - ta) = 0$.	The closest point is the foot of the perpendicular: the error $b - p$ must be orthogonal to the line. If it were not, sliding p along the line would shorten it.
$a^\top b - t a^\top a = 0$, so $t = \frac{a^\top b}{a^\top a}$.	Distribute $a^\top$; $a^\top a = \ a\ ^2 > 0$, so it can be divided by.
$p = \frac{a^\top b}{a^\top a} a = \frac{a(a^\top b)}{a^\top a} = \frac{aa^\top}{a^\top a} b$.	$a^\top b$ is a number, so it can move to the other side of a ; then regroup $a(a^\top b) = (aa^\top)b$.
$p = \frac{a^\top b}{a^\top a} a$, $P = \frac{aa^\top}{a^\top a}$	P is $n \times n$: a column times a row, divided by a number. Check: $a = (1, 0)^\top$ gives $P = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, which keeps the first coordinate and zeroes the second.

Problem 2

Show that P from Problem 1 satisfies $P^2 = P$ and $P^\top = P$, has rank 1, and has eigenvalues 1 and 0. What are the eigenvectors?

$P^2 = \frac{aa^\top aa^\top}{(a^\top a)^2} = \frac{a(a^\top a)a^\top}{(a^\top a)^2} = \frac{aa^\top}{a^\top a} = P$.	The middle $a^\top a$ is a number and cancels one factor of the denominator. Geometrically, projecting a point already on the line leaves it where it is.
$P^\top = \frac{(aa^\top)^\top}{a^\top a} = \frac{(a^\top)^\top a^\top}{a^\top a} = \frac{aa^\top}{a^\top a} = P$.	$(XY)^\top = Y^\top X^\top$, and $(a^\top)^\top = a$.

Column j of P is $\frac{a_j}{a^\top a} a$.	Column j of $aa^\top$ is a times the j -th entry of $a^\top$. Every column is a multiple of $a \neq 0$, so the column space is the line through a and the rank is 1.
$Pa = \frac{a(a^\top a)}{a^\top a} = a$.	So a is an eigenvector with eigenvalue 1.
For $x \perp a$: $Px = \frac{a(a^\top x)}{a^\top a} = 0 = 0 \cdot x$.	$a^\top x = 0$ is what $x \perp a$ means. So every nonzero x orthogonal to a is an eigenvector with eigenvalue 0.
The vectors orthogonal to a form a subspace of dimension $n - 1$, so $\lambda = 0$ has $n - 1$ independent eigenvectors; with a that makes n , so there are no other eigenvalues.	They are the solutions of one nonzero equation $a^\top x = 0$ in n unknowns. n independent eigenvectors account for all n eigenvalues, counted with multiplicity.
$P^2 = P$ and $P^\top = P$ by direct multiplication; rank 1 (every column is a multiple of a); $Pa = a$ so $\lambda = 1$ with eigenvector a, and $Px = 0$ for every $x \perp a$, so $\lambda = 0$ with multiplicity $n - 1$	In $\mathbb{R}^2$ with $a = (1, 1)^\top$: $P = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $P(1, 1)^\top = (1, 1)^\top$ and $P(1, -1)^\top = 0$. P is symmetric, so its eigenvectors for 1 and 0 are orthogonal, as on the previous page.

Problem 3

For A with independent columns, find the matrix that projects onto $\text{col}(A)$, and show the residual $b - Pb$ is orthogonal to every column of A .

$Pb = A\hat{x}$ for some $\hat{x} \in \mathbb{R}^k$.	Pb lies in $\text{col}(A)$, and every point there is A times some vector.
---	--

$b - A\hat{x} \perp \text{col}(A)$, that is $A^\top(b - A\hat{x}) = 0$.	As in Problem 1, the closest point is where the error is perpendicular to the whole subspace. Orthogonal to every column of A means every entry of $A^\top(b - A\hat{x})$, one column dotted with the error per entry, is 0.
$A^\top A\hat{x} = A^\top b$, so $\hat{x} = (A^\top A)^{-1}A^\top b$.	These are the normal equations, the same ones the matrix-calculus page reached by setting the gradient to 0. $A^\top A$ is invertible because the columns of A are independent.
$Pb = A\hat{x} = A(A^\top A)^{-1}A^\top b$ for every b , so $P = A(A^\top A)^{-1}A^\top$.	Read the matrix off. With one column a , $A^\top A = a^\top a$ is a number and this is Problem 1's P .
$P^2 = A(A^\top A)^{-1}(A^\top A)(A^\top A)^{-1}A^\top = A(A^\top A)^{-1}A^\top = P$.	The inner $(A^\top A)(A^\top A)^{-1}$ is I .
$P^\top = A((A^\top A)^{-1})^\top A^\top = A(A^\top A)^{-1}A^\top = P$.	Transpose reverses the product; the transpose of an inverse is the inverse of the transpose, and $(A^\top A)^\top = A^\top A$.
$P = A(A^\top A)^{-1}A^\top$; $A^\top(b - Pb) = A^\top b - A^\top b = 0$	$A^\top Pb = (A^\top A)(A^\top A)^{-1}A^\top b = A^\top b$. Each entry of $A^\top(b - Pb)$ is one column of A dotted with the residual, so all of them are 0.

Problem 4

Fit the line $y = c + mx$ to the points $(0, 1)$, $(1, 2)$, $(2, 2)$ by least squares: set up A and b , form the normal equations and solve them.

Each point (x, y) asks for $c + mx = y$. The three equations, in the unknowns c and m , are

$$A(c, m)^\top = b \text{ with } A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \text{ and } b = (1, 2, 2)^\top.$$

Row i of A is $(1, x_i)$ and entry i of b is y_i .

c is multiplied by 1 in every equation and m by the point's x . Three equations, two unknowns, and the points are not on one line, so there is no exact solution.

$A^T A = \begin{bmatrix} 1+1+1 & 0+1+2 \\ 0+1+2 & 0+1+4 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}.$	Entry (i, j) is column i of A dotted with column j .
$A^T b = (1+2+2, 0+2+4)^T = (5, 6)^T.$	Each column of A dotted with b .
$3c + 3m = 5$ and $3c + 5m = 6.$	The normal equations $A^T A (c, m)^T = A^T b$ written out.
$2m = 1$, so $m = \frac{1}{2}$; then $3c = 5 - \frac{3}{2} = \frac{7}{2}$, so $c = \frac{7}{6}.$	Subtract the first equation from the second to remove c , then substitute back.
$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, b = (1, 2, 2)^T; A^T A = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}, A^T b = (5, 6)^T; c = \frac{7}{6}, m = \frac{1}{2}$	Check in the first equation: $3 \cdot \frac{7}{6} + 3 \cdot \frac{1}{2} = \frac{7}{2} + \frac{3}{2} = 5$. The fitted line $y = \frac{7}{6} + \frac{1}{2}x$ passes at heights $\frac{7}{6}, \frac{5}{3}, \frac{13}{6}$ over $x = 0, 1, 2$.

Problem 5

Compute the residual $r = b - A\hat{x}$ of Problem 4 and verify $A^T r = 0$.

$\hat{x} = (\frac{7}{6}, \frac{1}{2})^T$ and $A\hat{x} = (\frac{7}{6}, \frac{7}{6} + \frac{1}{2}, \frac{7}{6} + 1)^T = (\frac{7}{6}, \frac{5}{3}, \frac{13}{6})^T.$	Row i of A is $(1, x_i)$, so $A\hat{x}$ lists the line's heights at the three x values.
$r = (1 - \frac{7}{6}, 2 - \frac{5}{3}, 2 - \frac{13}{6})^T = (-\frac{1}{6}, \frac{1}{3}, -\frac{1}{6})^T.$	Subtract entry by entry; each entry is how far the point is above the line.
First entry of $A^T r$: $-\frac{1}{6} + \frac{1}{3} - \frac{1}{6} = 0.$	The first column of A is all ones, so this is the sum of the residuals: a least-squares line with an intercept always balances its errors.
Second entry: $0 \cdot (-\frac{1}{6}) + 1 \cdot \frac{1}{3} + 2 \cdot (-\frac{1}{6}) = \frac{1}{3} - \frac{1}{3} = 0.$	The second column of A is the x values, so the residuals are also uncorrelated with x .

$$r = \left(-\frac{1}{6}, \frac{1}{3}, -\frac{1}{6}\right)^\top \text{ and } A^\top r = (0, 0)^\top$$

This is Problem 3's orthogonality on actual numbers: r is orthogonal to both columns, so to all of $\text{col}(A)$. The squared error is $\|r\|^2 = \frac{1}{36} + \frac{1}{9} + \frac{1}{36} = \frac{1}{6}$, and no other line does better.

Problem 6

Compute the SVD of $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$ by hand: singular values, V , then U .

$A^\top A = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}$, with eigenvalues 45 and 5 and eigenvectors $(1, 1)^\top$ and $(1, -1)^\top$.

The previous page computed all of this (its Problem 10). $A = U\Sigma V^\top$ gives $A^\top A = V\Sigma U^\top U\Sigma V^\top = V\Sigma^2 V^\top$, so the eigenvectors of $A^\top A$ are the v_i and its eigenvalues are the σ_i^2 .

$$\sigma_1 = \sqrt{45} = 3\sqrt{5} \text{ and } \sigma_2 = \sqrt{5}.$$

Square roots of the eigenvalues, largest first.

$$v_1 = \frac{1}{\sqrt{2}}(1, 1)^\top, v_2 = \frac{1}{\sqrt{2}}(1, -1)^\top, \text{ so } V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

The columns of V must be unit vectors, so divide each eigenvector by its length $\sqrt{2}$, in the order of the singular values.

$$Av_1 = \frac{1}{\sqrt{2}}(3, 4+5)^\top = \frac{1}{\sqrt{2}}(3, 9)^\top, \text{ and } u_1 = \frac{Av_1}{\sigma_1} = \frac{(3, 9)^\top}{\sqrt{2} \cdot 3\sqrt{5}} = \frac{1}{\sqrt{10}}(1, 3)^\top.$$

$Av_i = \sigma_i u_i$ (Before you start), so $u_i = Av_i/\sigma_i$. Computing U this way keeps its signs matched to V .

$$Av_2 = \frac{1}{\sqrt{2}}(3, 4-5)^\top = \frac{1}{\sqrt{2}}(3, -1)^\top, \text{ and } u_2 = \frac{(3, -1)^\top}{\sqrt{2} \cdot \sqrt{5}} = \frac{1}{\sqrt{10}}(3, -1)^\top.$$

Same rule with $\sigma_2 = \sqrt{5}$.

$$u_1^\top u_2 = \frac{1}{10}(3-3) = 0 \text{ and } u_1^\top u_1 = u_2^\top u_2 = \frac{1+9}{10} = 1.$$

They come out orthonormal automatically: $u_i^\top u_j = v_i^\top A^\top A v_j / (\sigma_i \sigma_j) = \sigma_j^2 v_i^\top v_j / (\sigma_i \sigma_j)$, which is 0 for $i \neq j$ and 1 for $i = j$.

$$U\Sigma = \frac{1}{\sqrt{10}} \begin{bmatrix} 3\sqrt{5} & 3\sqrt{5} \\ 9\sqrt{5} & -\sqrt{5} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 3 & 3 \\ 9 & -1 \end{bmatrix}.$$

Multiplying by a diagonal matrix on the right scales the columns of U by $3\sqrt{5}$ and $\sqrt{5}$; $\sqrt{5}/\sqrt{10} = 1/\sqrt{2}$.

$$U\Sigma V^\top = \frac{1}{2} \begin{bmatrix} 3 & 3 \\ 9 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 6 & 0 \\ 8 & 10 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}.$$

$V^\top = V$ here, and the two factors of $\frac{1}{\sqrt{2}}$ make $\frac{1}{2}$. Multiplying back is the check that nothing slipped.

$$\sigma_1 = 3\sqrt{5}, \sigma_2 = \sqrt{5}; V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix};$$

$$u_i = Av_i/\sigma_i \text{ gives } U = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 3 \\ 3 & -1 \end{bmatrix};$$

check $U\Sigma V^\top = A$

When the singular values are distinct and nonzero, as here, the SVD is unique up to signs: replacing v_i by $-v_i$ flips u_i too, and the product is unchanged.

Problem 7

A 3×2 matrix has singular values $\sigma_1 > 0$ and $\sigma_2 = 0$. Describe its rank, column space and null space in terms of u_1 and v_2 .

$$A = U\Sigma V^\top = \sigma_1 u_1 v_1^\top + \sigma_2 u_2 v_2^\top = \sigma_1 u_1 v_1^\top.$$

Multiply out column by row: $U\Sigma$ has columns $\sigma_i u_i$, and a product of matrices is the sum of each column of the first times the matching row of the second. The σ_2 term is 0.

$$Ax = \sigma_1 u_1 (v_1^\top x) \text{ for every } x \in \mathbb{R}^2.$$

$v_1^\top x$ is a number, so every output is a multiple of u_1 .

$$\text{col}(A) = \text{span}\{u_1\}, \text{ so the rank is 1.}$$

Step 2 puts every Ax on the line through u_1 , and $x = v_1$ gives $Av_1 = \sigma_1 u_1 \neq 0$, so the whole line is reached. The rank is the dimension of the column space.

$$Ax = 0 \text{ exactly when } v_1^\top x = 0, \text{ that is when } x \text{ is a multiple of } v_2.$$

$\sigma_1 u_1 \neq 0$, so the product in step 2 is 0 only when the number $v_1^\top x$ is. In $\mathbb{R}^2$ the vectors orthogonal to the unit vector v_1 are the multiples of v_2 , since V is orthogonal.

rank 1; $\text{col}(A) = \text{span}\{u_1\}$; **null space**
 $= \text{span}\{v_2\}$, **because** $Av_2 = \sigma_2 u_2 = 0$

In general the rank is the number of nonzero singular values, the u_i that go with them span the column space, and the v_i that go with zero singular values span the null space. The rank count agrees: 2 columns = 1 (rank) + 1 (null space dimension).

Problem 8

Define $A^+ = V\Sigma^+U^\top$. Show that when A has independent columns, $A^+ = (A^\top A)^{-1}A^\top$, and compute A^+b for Problem 4.

Independent columns make every $\sigma_i > 0$, so Σ is invertible and $\Sigma^+ = \Sigma^{-1}$.

$\sigma_i = \|Av_i\|$, because $Av_i = \sigma_i u_i$ with $\|u_i\| = 1$; and $Av_i \neq 0$ since $v_i \neq 0$ and only $x = 0$ gives $Ax = 0$ when the columns are independent.

$$A^\top A = V\Sigma U^\top U \Sigma V^\top = V\Sigma^2 V^\top.$$

Substitute $A = U\Sigma V^\top$ and $A^\top = V\Sigma U^\top$ (Σ is diagonal, so $\Sigma^\top = \Sigma$); $U^\top U = I$ removes the middle.

$$(A^\top A)^{-1} = V\Sigma^{-2}V^\top.$$

Multiply to check: $V\Sigma^2 V^\top V\Sigma^{-2}V^\top = V\Sigma^2 \Sigma^{-2}V^\top = VV^\top = I$, using $V^\top V = I$ and then $VV^\top = I$.

$$(A^\top A)^{-1}A^\top = V\Sigma^{-2}V^\top V\Sigma U^\top = V\Sigma^{-2}\Sigma U^\top = V\Sigma^{-1}U^\top.$$

$V^\top V = I$ again, and diagonal matrices multiply entry by entry: $\sigma_i^{-2}\sigma_i = \sigma_i^{-1}$.

$$\text{For Problem 4, } (A^\top A)^{-1} = \frac{1}{6} \begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix}.$$

The 2×2 inverse of $\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}$: determinant $15 - 9 = 6$, swap the diagonal, negate the off-diagonal.

$$A^+b = (A^\top A)^{-1}A^\top b = \frac{1}{6} \begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \end{bmatrix} = \frac{1}{6}(25 - 18, -15 + 18)^\top = \left(\frac{7}{6}, \frac{1}{2}\right)^\top.$$

Step 4 lets us use $A^\top b = (5, 6)^\top$ from Problem 4 instead of computing an SVD of that A .

$$(A^\top A)^{-1} A^\top = V \Sigma^{-1} U^\top = A^+; A^+ b = \left(\frac{7}{6}, \frac{1}{2}\right)^\top$$

The same c and m as Problem 4: when the columns are independent, $A^+ b$ is the least-squares solution $\hat{x}$. A^+ is still defined when they are not, which is why robust solvers such as NumPy's `lstsq` go through the SVD; there $A^+ b$ is the least-squares solution of smallest norm.

Problem 9

For A of Problem 6, write down the best rank-one approximation A_1 and the size of the error $\|A - A_1\|_2$.

The Eckart–Young theorem: among all matrices of rank at most 1, $A_1 = \sigma_1 u_1 v_1^\top$ is closest to A in $\|\cdot\|_2$, and $\|A - A_1\|_2 = \sigma_2$.

Quoted here, not proved (the spectral-norm form is Mirsky's). In $\|\cdot\|_2$ it is *a* closest matrix, not the only one; in the Frobenius norm it is the unique closest when $\sigma_1 > \sigma_2$. The same holds for rank j , with $A_j = \sum_{i \leq j} \sigma_i u_i v_i^\top$ and error σ_{j+1} .

$$A_1 = 3\sqrt{5} \cdot \frac{1}{\sqrt{10}}(1, 3)^\top \cdot \frac{1}{\sqrt{2}}(1, 1) = \frac{3\sqrt{5}}{\sqrt{20}} \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix} = \frac{3}{2} \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix}.$$

u_1, v_1 and σ_1 from Problem 6; $v_1^\top = \frac{1}{\sqrt{2}}(1, 1)$ is a row, and a column times a row gives the matrix of products. $\sqrt{10}\sqrt{2} = \sqrt{20} = 2\sqrt{5}$.

$$A - A_1 = \begin{bmatrix} 3 - \frac{3}{2} & 0 - \frac{3}{2} \\ 4 - \frac{9}{2} & 5 - \frac{9}{2} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix}.$$

Subtract entry by entry.

$$\frac{1}{2} \begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix} = \sqrt{5} \cdot \frac{1}{\sqrt{10}}(3, -1)^\top \cdot \frac{1}{\sqrt{2}}(1, -1) = \sigma_2 u_2 v_2^\top.$$

$A = \sigma_1 u_1 v_1^\top + \sigma_2 u_2 v_2^\top$ (Problem 7, step 1), so the error is exactly the dropped term; $\sqrt{5}/\sqrt{20} = \frac{1}{2}$.

$\|\sigma_2 u_2 v_2^\top x\| = \sigma_2 |v_2^\top x| \leq \sigma_2$ for unit x , with equality at $x = v_2$.

$\|u_2\| = 1$, and $|v_2^\top x| \leq \|v_2\| \|x\| = 1$ with equality when x points along v_2 ; so the largest stretch is σ_2 .

$$A_1 = \sigma_1 u_1 v_1^\top \text{ and } \|A - A_1\|_2 = \sigma_2 = \sqrt{5}$$

Here $A_1 = \frac{3}{2} \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix}$. The error is the largest singular value thrown away, so a matrix whose singular values fall off fast is well approximated by few terms: the idea behind low-rank compression and PCA.

Problem 10

Show $I - P$ projects onto the orthogonal complement of $\text{col}(A)$, and that $\|b\|^2 = \|Pb\|^2 + \|(I - P)b\|^2$.

$P = A(A^\top A)^{-1}A^\top$ is the projection of Problem 3, with $P^2 = P$ and $P^\top = P$.

$$(I - P)^2 = I - 2P + P^2 = I - 2P + P = I - P, \text{ and } (I - P)^\top = I - P.$$

Expand, then $P^2 = P$; the transpose of I is I and $P^\top = P$. So $I - P$ is a projection matrix of the same kind as P .

$$A^\top(I - P) = A^\top - A^\top A(A^\top A)^{-1}A^\top = A^\top - A^\top = 0.$$

So every $(I - P)b$ is orthogonal to every column of A : $I - P$ lands in $\text{col}(A)^\perp$.

$$\text{For } w \in \text{col}(A)^\perp: Pw = A(A^\top A)^{-1}(A^\top w) = 0, \text{ so } (I - P)w = w.$$

$A^\top w = 0$ is what orthogonal to every column means. $I - P$ leaves that subspace fixed, so with step 2 its range is exactly $\text{col}(A)^\perp$. And $b - (I - P)b = Pb$ lies in $\text{col}(A)$, orthogonal to that subspace, so $(I - P)b$ is the foot of the perpendicular from b , as in Problem 3; it is the residual $b - Pb$.

$$b = Pb + (I - P)b, \text{ so } \|b\|^2 = \|Pb\|^2 + 2(Pb)^\top(I - P)b + \|(I - P)b\|^2.$$

$$\text{Expand } \|w + z\|^2 = (w + z)^\top(w + z) = \|w\|^2 + 2w^\top z + \|z\|^2.$$

$$(Pb)^\top(I - P)b = b^\top P^\top(I - P)b = b^\top(P - P^2)b = 0.$$

$P^\top = P$ and $P^2 = P$, so $P^\top(I - P) = P - P^2$ is the zero matrix. The cross term vanishes.

$I - P$ is the projection onto $\text{col}(A)^\perp$,
and $\|b\|^2 = \|Pb\|^2 + \|(I - P)b\|^2$

Pythagoras: b splits into perpendicular pieces. For Problem 4, $\|b\|^2 = 9$, $\|r\|^2 = \frac{1}{6}$ (Problem 5) and $\|A\hat{x}\|^2 = \frac{49}{36} + \frac{100}{36} + \frac{169}{36} = \frac{53}{6}$; indeed $\frac{53}{6} + \frac{1}{6} = 9$.

Where this goes wrong

1. Projecting with $AA^\top$

$aa^\top$ is the top of the projection onto a line, and $AA^\top$ looks like its many-column version.

Project $b = (1, 2, 2)^\top$ onto $\text{col}(A)$ for the A of Problem 4

Right so far: the answer should be $A\hat{x} = (\frac{7}{6}, \frac{5}{3}, \frac{13}{6})^\top$.

“For a line it is $aa^\top$ over something; for a matrix, $AA^\top$.”

The half-memory that causes the mistake: the denominator of Problem 1, and what it becomes for many columns, is dropped.

$$P = AA^\top \times$$

Right only when the columns of A are orthonormal, so that $A^\top A = I$; in general $(A^\top A)^{-1}$ undoes the columns' lengths and angles. Here $AA^\top b = (5, 11, 17)^\top$, nowhere near b , and $AA^\top$ is not a projection: its top-left entry is 1, while that of $(AA^\top)^2$ is 3. The test $P^2 = P$ catches it.

2. Solving $Ax = b$ when it has no solution

A system of equations invites row reduction, and row reduction is the right tool only when a solution exists.

$c = 1$, $c + m = 2$, $c + 2m = 2$ for the points of Problem 4

Right so far: these are the three equations $A(c, m)^\top = b$.

“Row-reduce $[A \mid b]$ and solve.”

The step that causes the mistake: it looks for an exact solution.

The first two equations give $c = 1, m = 1$; the third then says $3 = 2$, so “no line fits and there is no answer.” ✗

Three equations, two unknowns: the inconsistency is expected, since the three points are not on one line. Least squares solves $A^\top A \hat{x} = A^\top b$ instead, which always has a solution, and gives $c = \frac{7}{6}, m = \frac{1}{2}$. The line through the first two points, $c = 1, m = 1$, misses the third by 1 and has squared error 1, against $\frac{1}{6}$ for the fit.

3. Singular values as eigenvalues of A

For symmetric matrices with nonnegative eigenvalues the two lists agree, and the habit carries over.

$A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$ from Problem 6

Right so far.

“ A is triangular, so its eigenvalues are 3 and 5 on the diagonal.”

True, and the step that sets up the mistake: an easy list of numbers is at hand.

$\sigma_i = \lambda_i(A)$, so $\sigma_1 = 5$ and $\sigma_2 = 3$ ✗

Singular values are the square roots of the eigenvalues of $A^\top A$: for Problem 6 the eigenvalues of A are 3 and 5 and the singular values are $\sqrt{5}$ and $3\sqrt{5}$. The check: $\sigma_1 = \|A\|_2$ is the largest stretch, and $\|Av_1\| = \|\frac{1}{\sqrt{2}}(3, 9)^\top\| = \sqrt{45} = 3\sqrt{5} > 5$.

4. U from the eigenvectors of $A^\top A$

$A^\top A$ is the matrix the singular values came from, so it is tempting to take both U and V from it.

$A^\top A$ for Problem 6 has unit eigenvectors $\frac{1}{\sqrt{2}}(1, 1)^\top$ and $\frac{1}{\sqrt{2}}(1, -1)^\top$

Right so far.

“The SVD needs two orthogonal matrices, and here are two orthogonal eigenvectors.”

The step that causes the mistake: $A^\top A$ has already supplied V .

U is built from the eigenvectors of $A^\top A$,
that is $U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ ✗

Those are V ; U comes from $AA^\top$ or
from $u_i = Av_i/\sigma_i$. With this U , $U\Sigma V^\top =$
 $\sqrt{5} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, a symmetric matrix, and A is
not symmetric. Multiplying back catches
it.

5. Residual orthogonal to b

“The residual is orthogonal” is remembered without the thing it is orthogonal to.

$r = (-\frac{1}{6}, \frac{1}{3}, -\frac{1}{6})^\top$ for Problem 4

Right so far: Problem 5.

“Check the fit by confirming the residual
is orthogonal.”

The step that causes the mistake:
orthogonal to what is left unsaid, and b
is the vector at hand.

Check $b^\top r = 0$: $-\frac{1}{6} + \frac{2}{3} - \frac{1}{3} = \frac{1}{6} \neq 0$, so
“the fit is wrong” ✗

The residual is orthogonal to the *column
space*, not to b . Since $b = A\hat{x} + r$ and
 $A^\top r = 0$, $b^\top r = \hat{x}^\top A^\top r + r^\top r = \|r\|^2 > 0$
whenever the fit is not exact; here $\frac{1}{6} =$
 $\|r\|^2$, exactly as it should be. The right
test is $A^\top r = 0$.

eulerpad.app/practice/least-squares-projections-and-the-svd