Inverse Gaussian Elimination

Sure 😊 Let's do it in a more compact way.

We want A^{-1} for

$$A = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Start with the augmented matrix $[A \mid I]$:

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$$

Main row operations (listed without showing every intermediate matrix)

1. Make zeros under the first pivot (column 1):

$$R_2 \leftarrow R_2 - R_1, \quad R_3 \leftarrow R_3 - R_1, \quad R_4 \leftarrow R_4 - R_1$$

2. Use pivot in column 2 (R_2) to clear entries below:

$$R_3 \leftarrow R_3 - 2R_2, \quad R_4 \leftarrow R_4 - R_2$$

3. Swap rows R_3 and R_4 (to get a 1 as pivot in column 3), then clear below:

$$R_3 \leftrightarrow R_4, \quad R_4 \leftarrow R_4 - 2R_3$$

4. Make the pivot in row 4 equal to 1, then clear above it in column 4.
5. Use the pivots in columns 3, 2, 1 to clear above them (back-substitution style) so that the left block becomes I_4.

After these operations, the matrix becomes

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & -1 & 2 & -2 & 2 \\ 0 & 1 & 0 & 0 & 1 & -1 & 2 & -2 \\ 0 & 0 & 1 & 0 & 1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 & 2 \end{array}\right]$$

So the inverse is just the right block:

$$\boxed{A^{-1} = \begin{bmatrix} -1 & 2 & -2 & 2 \\ 1 & -1 & 2 & -2 \\ 1 & -1 & 1 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix}}$$

If you want, I can write ONLY the row operations in one list with their results, no explanations.

Reduced Row PREF

Step 1 — Identify pivot and free columns

Matrix A in RREF:

$$A = \begin{bmatrix} 1 & 3 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 9 \\ 0 & 0 & 0 & 1 & -4 \end{bmatrix}$$

Pivot columns: 1, 3, 4

Free columns: 2, 5

So the null space will have 2 parameters.

Step 2 — Apply the Minus-1 Trick

Create a 5×5 matrix by inserting −1 on the diagonal where pivot columns are missing (columns 2 and 5):

$$\bar{A} = \begin{bmatrix} 1 & 3 & 0 & 0 & 3 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 9 \\ 0 & 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Step 3 — Read solution vectors

Take the columns that have −1 **on the diagonal:**

Column 2 → first solution vector

$$v_1 = \begin{bmatrix} 3 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Column 5 → second solution vector

$$v_2 = \begin{bmatrix} 3 \\ 0 \\ 9 \\ -4 \\ -1 \end{bmatrix}$$

Final Answer (in few steps)

$$x = \lambda_1 \begin{bmatrix} 3 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 3 \\ 0 \\ 9 \\ -4 \\ -1 \end{bmatrix}, \quad \lambda_1, \lambda_2 \in \mathbb{R}.$$

Augment and PREF

$x_1 + 2x_2 - x_3 + 3x_4 = 5$
$2x_1 + 4x_2 + x_3 + x_4 = 8$
$3x_1 - 6x_2 + 2x_3 + (a+1)x_4 = -10-$

(A) Write the augmented matrix.

(B) Reduce it to REF.

(C) Determine for which values of a the system is:

1. inconsistent
2. has unique solution
3. has infinitely many solutions

(D) For solvable cases, find the particular solution.

(E) Find the general solution (parametric vector form).

(F) Solve the homogeneous system $Ax = 0$.

(G) Use the minus-1 trick to find the null space.

(H) Identify pivot and free variables.

SOLUTION — Problem 1

(A) Augmented matrix

$$\begin{bmatrix} 1 & 2 & -1 & 3 & 5 \\ 2 & 4 & 1 & 1 & 8 \\ -3 & -6 & 2 & a+1 & -10 \end{bmatrix}$$

(B) REF

1. $R2 \to R2 - 2R1$
2. $R3 \to R3 + 3R1$

$$\begin{bmatrix} 1 & 2 & -1 & 3 & 5 \\ 0 & 0 & 3 & -5 & -2 \\ 0 & 0 & -1 & a+10 & 5 \end{bmatrix}$$

PR

Giv

(A)

(B)

(C)

(D)

(E)

(A)

Piv
Fre

(B)

Equ

So

:OBLEM 2

ven matrix in RREF:

$$A = \begin{bmatrix} 1 & 0 & 3 & 1 & 0 \\ 0 & 1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

) Identify pivot / free variables

) Solve Ax = 0

) Use minus-1 trick to find null space basis

) Write general solution

) Compute inverse of 2×2 matrix using Gaussian elimination:

$$B = \begin{bmatrix} 4 & 7 \\ 2 & 3 \end{bmatrix}$$

) Pivot / free

/ots in columns: 1, 2, 4, 5
:e variable: x3 only

) Solve Ax = 0

uations:

1. $x_1 + 3x_3 + x_4 = 0$
2. $x_2 - 2x_3 = 0$
3. $x_4 = 0$
4. $x_5 = 0$

:

- x4 = 0
- x5 = 0
- x2 = 2x3

(E) Inverse

Form:

1. R2 ·
2. R2 ·
3. R1 ·
4. R1 ·

Thus:

e of B

$$B = \begin{bmatrix} 4 & 7 \\ 2 & 3 \end{bmatrix}$$

$$[B \mid I] = \begin{bmatrix} 4 & 7 & | & 1 & 0 \\ 2 & 3 & | & 0 & 1 \end{bmatrix}$$

→ R2 − (1/2)R1:

$$\begin{bmatrix} 4 & 7 & | & 1 & 0 \\ 0 & -0.5 & | & -0.5 & 1 \end{bmatrix}$$

→ R2 / -0.5:

$$\begin{bmatrix} 4 & 7 & | & 1 & 0 \\ 0 & 1 & | & 1 & -2 \end{bmatrix}$$

→ R1 − 7•R2:

$$\begin{bmatrix} 4 & 0 & | & -6 & 14 \\ 0 & 1 & | & 1 & -2 \end{bmatrix}$$

→ R1/4:

$$[B \mid B^{-1}] = \begin{bmatrix} 1 & 0 & | & -\frac{3}{2} & \frac{7}{2} \\ 0 & 1 & | & 1 & -2 \end{bmatrix}$$

$$B^{-1} = \begin{bmatrix} -\frac{3}{2} & \frac{7}{2} \\ 1 & -2 \end{bmatrix}$$

Angles and Orthogonality

For vectors x, y,

$$\cos\omega = \frac{\langle x, y\rangle}{\sqrt{\langle x, x\rangle \langle y, y\rangle}} = \frac{x^\top y}{\sqrt{x^\top x \, y^\top y}}$$

2 Compute the dot products

a) $x^\top y$

$$x^\top y = \begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix} = 1\cdot 1 + 1\cdot 2 = 1 + 2 = 3$$

b) $x^\top x$ (norm squared of x)

$$x^\top x = \begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} = 1\cdot 1 + 1\cdot 1 = 1 + 1 = 2$$

c) $y^\top y$ (norm squared of y)

$$y^\top y = \begin{bmatrix}1 & 2\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix} = 1\cdot 1 + 2\cdot 2 = 1 + 4 = 5$$

3 Plug into the formula

$$\cos\omega = \frac{x^\top y}{\sqrt{x^\top x \, y^\top y}} = \frac{3}{\sqrt{2\cdot 5}} = \frac{3}{\sqrt{10}}$$

So

$$\omega = \arccos\left(\frac{3}{\sqrt{10}}\right)$$

4 Numerical value

$$\frac{3}{\sqrt{10}} \approx 0.9487$$

$$\omega \approx \arccos(0.9487) \approx 0.32 \text{ radians} \approx 18^\circ$$

P

Comput

Since th

So the a

This

P

The inn

where

Step 1

Step 2

So:

+

Orthogonal Vectors

'ART 1 — Using the Standard Dot Product

te:

$$x^\top y = [1 \quad 1]\begin{bmatrix}-1\\1\end{bmatrix} = 1(-1) + 1(1) = -1 + 1 = 0$$

ıe dot product is 0, we know:

$$x \perp y$$

angle between them is:

$$\omega = 90^\circ$$

; matches the picture (the arrows look perpendicular).

'ART 2 — Using the New Inner Product

er product is defined as:

$$\langle x, y \rangle_M = x^\top M y$$

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

---

l: Compute My

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

2: Compute the inner product $x^\top My$

$$x^\top (My) = [1 \quad 1]\begin{bmatrix}-2\\1\end{bmatrix} = 1(-2) + 1(1) = -2 + 1 = -1$$

Ask anything

Use

We a

Now

Con

Con

We a

Plug

Con

PART 3 — Compute the New Angle (optional but useful)

the generalized angle formula:

$$\cos\omega = \frac{x^\top M y}{\sqrt{x^\top M x \;\; y^\top M y}}$$

already have:

$$x^\top M y = -1$$

/ compute the norms.

---

npute $x^\top M x$

$$Mx = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$x^\top M x = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 2 + 1 = 3$$

---

npute $y^\top M y$

already found:

$$My = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

$$y^\top M y = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \end{bmatrix} = (-1)(-2) + 1 \cdot 1 = 2 + 1 = 3$$

---

g into the angle formula

$$\cos\omega = \frac{-1}{\sqrt{3 \cdot 3}} = \frac{-1}{3}$$

$$\omega = \arccos\left(-\frac{1}{3}\right) \approx 1.91 \text{ rad}$$

vert to degrees:

$$1.91 \text{ rad} \approx 109.5^\circ$$

- Ask anything

Eigenvalues and Eigenvec

We're given the matrix

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

We want:

1. Eigenvalues λ
2. Eigenvectors for each eigenvalue
3. Eigenspaces (the set of all eigenvectors for a given λ)

Step 1 – Characteristic Polynomial

Eigenvalues come from:

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

First compute $A - \lambda I$:

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

Now compute its determinant:

$$p_A(\lambda) = \det(A - \lambda I) = (4 - \lambda)(3 - \lambda) - 2 \cdot 1.$$

Multiply out:

$$(4 - \lambda)(3 - \lambda) = 12 - 4\lambda - 3\lambda + \lambda^2 = \lambda^2 - 7\lambda + 12.$$

Subtract 2:

$$p_A(\lambda) = \lambda^2 - 7\lambda + 12 - 2 = \lambda^2 - 7\lambda + 10.$$

So the characteristic polynomial is

$$p_A(\lambda) = \lambda^2 - 7\lambda + 10.$$

Step 2 – Eigenvalues

Solve

$$\lambda^2 - 7\lambda + 10 = 0.$$

Factor:

$$\lambda^2 - 7\lambda + 10 = (\lambda - 2)(\lambda - 5).$$

So the eigenvalues are

$$\boxed{\lambda_1 = 2, \quad \lambda_2 = 5}.$$

ctors

For each eigenvalue λ, solve

$$(A - \lambda I)x = 0$$

for a nonzero vector x.

(a) Eigenvectors for $\lambda = 5$

Compute $A - 5I$:

$$A - 5I = \begin{bmatrix} 4-5 & 2 \\ 1 & 3-5 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Solve

$$\begin{bmatrix} -1 & 2 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This gives equations:

$$-x_1 + 2x_2 = 0 \quad \text{and} \quad x_1 - 2x_2$$

They are the same equation: $x_1 = 2x_2$.

Let $x_2 = 1$ (any nonzero number works); then $x_1 = 2$.

So one eigenvector is

$$v_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

All eigenvectors for $\lambda = 5$ are scalar multiples of this:

$$E_5 = \left\{ \alpha \begin{bmatrix} 2 \\ 1 \end{bmatrix} : \alpha \in \mathbb{R} \right\} = \text{span}\left\{ \right[$$

That's the **eigenspace for** $\lambda = 5$.

(b) Eigenvectors for $\lambda = 2$

Compute $A - 2I$:

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

Solve

$$\begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Equations:

$$2x_1 + 2x_2 = 0, \qquad x_1 + x_2 =$$

They're the same equation: $x_1 = -x_2$.

Let $x_2 = -1$; then $x_1 = 1$.

So an eigenvector is

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

All eigenvectors for $\lambda = 2$ are multiples of this:

$$E_2 = \left\{ \alpha \begin{bmatrix} 1 \\ -1 \end{bmatrix} : \alpha \in \mathbb{R} \right\} = \text{span} \left\{ \left[\right. \right.$$

This is the **eigenspace for $\lambda = 2$**.

Final Result

- **Eigenvalues:**

$$\boxed{\lambda_1 = 2, \quad \lambda_2 = 5}$$

-

$\begin{matrix} 2 \\ -2 \end{matrix}\Big] .$

$= 0.$

$\begin{bmatrix} 2 \\ 1 \end{bmatrix} \Big\} .$

$\begin{matrix} 2 \\ 1 \end{matrix}\Big] .$

$0.$

4.2

Consider the following matrices:

$$A = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix},$$

---

(1) Eigenvalue equation & characteristi

1. (a) Write the **eigenvalue equatio**
2. (b) Derive the **characteristic poly**
3. (c) Find the **eigenvalues** of A and

---

(2) Eigenspaces, geometric multiplicity,

2. (a) For each eigenvalue $\lambda$of A, co

E

by solving $(A - \lambda I)x = 0$.

2. (b) For each λ, find a basis of E_λa $\dim\,(E_\lambda)$.
3. (c) Normalize the eigenvectors to
4. (d) Form the matrix

whose columns are the unit eigenvector

.

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix} \Big\} .$$

$B = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}, P = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$

---

c polynomial for A

on for A: $Ax = \lambda x$.

ynomial $p_A(\lambda) = \det(A - \lambda I)$.

l their **algebraic multiplicities**.

---

and eigenvectors of A

mpute the eigenspace

$E_\lambda = \ker(A - \lambda I)$

and its **geometric multiplicity**

o unit length and show that they are **orthonormal**.

$Q = [\, u_1 \; u_2 \,]$

rs of A, and the diagonal matrix

$\Lambda = \text{diag}(\lambda_1, \lambda_2)$.

Verify the s

---

(3) Determ

3. (a)

4. (b)

3. (c)
   Alo

---

(4) Identity

4. (a)
5. (b)
6. (c)

---

(5) Defecti

5. (a)
   mu
6. (b)
7. (c)

---

(6) Symme

6. (a)
7. (b)

spectral decomposition (spectral theorem for symmetric matrices):

$$A = Q\Lambda Q^{\top}.$$

ıinant, trace, and geometric interpretation for A

Compute $\det(A)$ and $\operatorname{tr}(A)$ directly from A.

Verify the theorems:

$$\det(A) = \prod_i \lambda_i, \operatorname{tr}(A) = \sum_i \lambda_i.$$

Give a geometric interpretation in $\mathbb{R}^2$:
ıng each eigenvector direction, by what factor does A stretch or compress the vector?

---

y matrix as a special case

Consider the identity matrix I_2. Find its characteristic polynomial and eigenvalues.

What is the eigenspace E_1 of I_2? What is its geometric multiplicity?

Explain why **every non-zero vector** in $\mathbb{R}^2$ is an eigenvector of I_2.

---

ve matrix and multiplicities: matrix B

Compute the characteristic polynomial of B and find its eigenvalues and **algebraic ıltiplicities**.

Find the eigenspace $E_2 = \ker(B - 2I)$ and its **geometric multiplicity**.

Decide whether B is **defective** or not (compare geometric vs algebraic multiplicity).

ɛtric positive semi-definite matrix $S = A^{\top}A$

Compute $S = A^{\top}A$ for matrix A.

Show that S is symmetric.

✅ Understanding Example 4.12 (Vectors and the SVD)

We are given the matrix

$$A = \begin{bmatrix} 1 & -0.8 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

This matrix maps vectors from $\mathbb{R}^2$ **to** $\mathbb{R}^3$.

We want to understand this mapping using the **Singular Value Decomposition** (SVD).

⭐ Step 1 — What is SVD?

SVD writes any matrix A as:

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

Where:

- $V^\top$ rotates vectors in the **input space** ($\mathbb{R}^2$).
- Σ stretches or shrinks vectors (scaling).
- U rotates vectors in the **output space** ($\mathbb{R}^3$).

So the mapping:

$$x \mapsto Ax$$

is broken into **three geometric operations**:

$$x \xrightarrow{V^\top} \text{rotate} \xrightarrow{\Sigma} \text{stretch} \xrightarrow{U} \text{rotate}$$

This is why SVD is so powerful:

It explains *exactly* how a matrix transforms vectors.

⭐ Step 2 — Look at the SVD of A

The book gives the decomposition:

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

Where

1. U (3×3 orthogonal matrix → rotation in $\mathbb{R}^3$)

$$U = \begin{bmatrix} -0.79 & 0 & -0.62 \\ 0.38 & -0.78 & -0.49 \\ -0.48 & -0.62 & 0.62 \end{bmatrix}$$

2. Σ (a 3×2 diagonal matrix → scaling)

$$\Sigma = \begin{bmatrix} 1.62 & 0 \\ 0 & 1.0 \\ 0 & 0 \end{bmatrix}$$

This tells us:

- First singular value = 1.62
- Second singular value = 1.00

These values tell how much the matrix **stretches** a rotated input vector.

3. $V^\top$ (2 × 2 rotation matrix — rotates the input grid)

$$V^\top = \begin{bmatrix} -0.78 & 0.62 \\ -0.62 & -0.78 \end{bmatrix}$$

⭐ Step 3 — How the SVD transforms the grid of vec

The example starts with a **2×2 square grid** of vectors in $\mathbb{R}^2$:

$$\mathcal{X} \subset [-1,1] \times [-1,1]$$

These vectors undergo three transformations:

---

(1) Apply $V^\top$ → Rotate the grid in $\mathbb{R}^2$

The original square grid becomes a **rotated square**.

This is shown in Figure 4.9 (not included here).

---

(2) Apply Σ → Stretch the grid in two perpendicular directions

- One axis is stretched by **1.62**
- The other axis is stretched by **1.0** (no change)

Now the rotated square becomes a **long ellipse/cylinder shape** in $\mathbb{R}^3$.

---

(3) Apply U → Rotate the stretched shape inside $\mathbb{R}^3$

This positions the final set of vectors in their correct 3D orientation.

This explains the final picture shown in the bottom-right of Figure 4.9.

⭐ What the example is teaching you

The key message:

SVD explains the geometry of any linear transformation:

1. Rotate the input
2. Stretch the input along orthogonal axes
3. Rotate into the output space

Even when mapping $\mathbb{R}^2 \to \mathbb{R}^3$, SVD still works and provides interpretation.

This is extremely helpful for:

- dimensionality reduction
- PCA

- image compression
- understanding transformations

⭐ Summary

- Start with a 2D grid of points.
- V^T rotates these points.
- Σ scales them along new orthogonal directions.
- U embeds and rotates them in 3D space.

Thus, SVD dissects the transformation A into pure geometric steps.

---

tors

4 5 ·

Exercise: Full SVD Analysis of a Non-Square Matrix

Consider the matrix

$$A = \begin{bmatrix} 3 & 1 \\ -1 & 3 \end{bmatrix}$$

This is a 2×3matrix with rank 2.

Perform all tasks below.

Task 1 — Compute the Full SVD of A

Step 1 — Compute $A^T A$

$$A^T A = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ -1 & 3 \end{bmatrix}$$

Step 2 — Eigenvalues of $A^T A$

Solve $\det(A^T A - \lambda I) = 0$ and obtain:

$$\lambda_1 = 12, \lambda_2 =$$

Step 3 — Singular Values

$$\sigma_i = \sqrt{\lambda_i}$$
$$\sigma_1 = 2\sqrt{3}, \sigma_2 = 2$$

Task 2 — Right-Singular Vectors

Compute eigenvectors of $A^T A$.
After normalization:

✅ Matrix we are working with

$$A^T A$$

To find λ, solve:

$$\det(A^T A$$

✅ Step 1 — Subtract λ from diag

$$A^T A - \lambda I =$$

✅ Step 2 — Take the determinan

$$\det \begin{bmatrix} 5-\lambda & 4 \\ 4 & 5-\lambda \end{bmatrix}$$

✅ Step 3 — Expand it

$$(5-\lambda)$$

Expand square:

$$(25 - 10\lambda$$

Simplify:

$$\lambda^2 - 1$$

✅ Step 4 — Solve the quadratic

$$\lambda^2 - 1$$

Use the formula:

$$10$$

$$\lambda = \frac{10}{}$$

Inside the square root:

$$100$$

$$\lambda =$$

$$\lambda =$$

:

$1 \quad 1]$
$3 \quad 1$

$\begin{matrix}1\\1\end{matrix}] = \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix}$

$8, \lambda_3 = 0$

$2\sqrt{2}, \sigma_3 = 0$

Thus

Task 3 — Build tl

Because A is $2 \times$

Task 4 — Compu

For u_1:

For u_2:

$= \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$

$A - \lambda I) = 0.$

onal

$= \begin{bmatrix} 5-\lambda & 4 \\ 4 & 5-\lambda \end{bmatrix}$

t

$= (5-\lambda)(5-\lambda) - 4 \cdot 4$

$)^2 - 16 = 0$

$+ \lambda^2) - 16 = 0$

$0\lambda + 9 = 0$

equation

$0\lambda + 9 = 0$

$$\frac{\pm\sqrt{100-36}}{2}$$

$$-36 = 64$$

$$\frac{10 \pm \sqrt{64}}{2}$$

$$= \frac{10 \pm 8}{2}$$

$$v_1 = \frac{1}{\sqrt{6}}\begin{bmatrix}1\\1\\2\end{bmatrix}, v_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\1\\0\end{bmatrix}, v_3 = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\-1\\1\end{bmatrix}$$

$$V = [v_1 \quad v_2 \quad v_3]$$

---

he Matrix Σ

3:

$$\Sigma = \begin{bmatrix}2\sqrt{3} & 0 & 0\\ 0 & 2\sqrt{2} & 0\end{bmatrix}$$

ıte the Left-Singular Vectors

$$u_i = \frac{1}{\sigma_i} A v_i$$

$$Av_1 = \frac{1}{\sqrt{6}}\begin{bmatrix}3+1+2\\-1+3+2\end{bmatrix} = \frac{1}{\sqrt{6}}\begin{bmatrix}6\\4\end{bmatrix} = \begin{bmatrix}\sqrt{6}\\ \frac{4}{\sqrt{6}}\end{bmatrix}$$

$$u_1 = \begin{bmatrix}\sqrt{\frac{1}{2}}\\ \frac{2}{3}\sqrt{\frac{1}{2}}\end{bmatrix}$$

Q2. SVD of $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$

1. $A^T A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$.
2. Eigenvalues: $\lambda_1 = 9, \ \lambda_2 = 1.$
   $\Rightarrow \sigma_1 = 3, \ \sigma_2 = 1.$
3. Eigenvectors (right singular vectors):
   - For $\lambda_1 = 9$: eigenvector $(1, 1)^T$.
     Normalized:
     $v_1 = \frac{1}{\sqrt{2}}(1, 1)^T$.
   - For $\lambda_2 = 1$: eigenvector $(-1, 1)^T$.
     Normalized:
     $v_2 = \frac{1}{\sqrt{2}}(-1, 1)^T$.

   $\Rightarrow$

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

4. Left singular vectors:

   $u_i = \frac{1}{\sigma_i} A v_i$.

   - $A v_1 = \frac{3\sqrt{2}}{2}(1, 1)^T \Rightarrow$
     $u_1 = \frac{1}{3} A v_1 = \frac{1}{\sqrt{2}}(1, 1)^T$.
   - $A v_2 = \frac{\sqrt{2}}{2}(1, -1)^T \Rightarrow$
     $u_2 = A v_2 = \frac{1}{\sqrt{2}}(1, -1)^T$.

   $\Rightarrow$

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

5. $\Sigma = \mathrm{diag}(3, 1)$.

SVD:

$$A = U\Sigma V^T.$$

(Because A is symmetric, here $U = V$.)

Q3. SVD of

1. $A^T A = \Big[$
2. Eigenvalue
   $\Rightarrow \sigma_1 = \surd$
3. Right singu
   - For λ_1
     $v_1 =$
   - For λ_2
     $v_2 =$

   $\Rightarrow$

$\begin{bmatrix} -1 \\ 1 \end{bmatrix}.$

4. Left singul
   - $Av_1 =$
     Norm

     $u_1 =$
   - $Av_2 =$
     Norm

     $u_2 =$

   To get a **fu**
   solve $\Rightarrow u_3$

$\begin{bmatrix} 1 \\ -1 \end{bmatrix}.$

   So

5. Σ (full, 3 >

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \quad (3 \times 2)$$

$$A^T A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

es: $\lambda_1 = 3, \ \lambda_2 = 1.$

$\sqrt{3}, \ \sigma_2 = 1.$

ular vectors:

$= 3$: eigenvector $(1,1)^T$,

$\frac{1}{\sqrt{2}}(1,1)^T.$

$= 1$: eigenvector $(-1,1)^T$, so

$\frac{1}{\sqrt{2}}(-1,1)^T.$

$$V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

ar vectors:

$= (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, \sqrt{2})^T.$

$= \sqrt{3} \Longrightarrow$

$$\frac{1}{\sqrt{3}} A v_1 = \begin{bmatrix} \sqrt{6}/6 \\ \sqrt{6}/6 \\ \sqrt{6}/3 \end{bmatrix}.$$

$= (-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 0)^T.$

$= 1 \Longrightarrow$

$$A v_2 = \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \\ 0 \end{bmatrix}.$$

ll $U \in \mathbb{R}^{3\times 3}$, add u_3 orthogonal to both:

$= \frac{1}{\sqrt{3}}(-1,-1,1)^T.$

$$U = [u_1 \ u_2 \ u_3].$$

< 2):

$$\Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ \downarrow & 1 \\ 0 & 0 \end{bmatrix}.$$

Example 4.12 (Vectors and the SVD)
Consider a mapping of a square grid of vectors $\mathcal{X} \in \mathbb{R}^2$ that fit in a box of size 2×2 centered at the origin. Using the standard basis, we map these vectors using

$$\boldsymbol{A} = \begin{bmatrix} 1 & -0.8 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top \tag{4.67a}$$

$$= \begin{bmatrix} -0.79 & 0 & -0.62 \\ 0.38 & -0.78 & -0.49 \\ -0.48 & -0.62 & 0.62 \end{bmatrix} \begin{bmatrix} 1.62 & 0 \\ 0 & 1.0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -0.78 & 0.62 \\ -0.62 & -0.78 \end{bmatrix}. \tag{4.67b}$$

We start with a set of vectors $\mathcal{X}$ (colored dots; see top-left panel of Figure 4.9) arranged in a grid. We then apply $\boldsymbol{V}^\top \in \mathbb{R}^{2\times 2}$, which rotates $\mathcal{X}$. The rotated vectors are shown in the bottom-left panel of Figure 4.9. We now map these vectors using the singular value matrix $\boldsymbol{\Sigma}$ to the codomain $\mathbb{R}^3$ (see the bottom-right panel in Figure 4.9). Note that all vectors lie in

1 What matrix A is doing

We have a linear map

$$\boldsymbol{A} : \mathbb{R}^2 \to \mathbb{R}^3, \qquad \boldsymbol{A} = \begin{bmatrix} 1 & -0.8 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

So any vector $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$ is mapped to

$$\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} 1 & -0.8 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - 0.8x_2 \\ x_2 \\ x_1 \end{bmatrix} \in \mathbb{R}^3.$$

The example says: imagine a **square grid** of points in $\mathbb{R}^2$ (like dots in a 2×2 box centered at the origin). We apply A to each point and see where they land in $\mathbb{R}^3$.

2 SVD of A

The SVD writes A as

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top,$$

where

- V is 2×2 orthogonal (rotation/reflection in the **input** space),
- Σ is 3×2 diagonal (singular values = scalings),
- U is 3×3 orthogonal (rotation/reflection in the **output** space).

In the example this is:

$$A = \underbrace{\begin{bmatrix} -0.79 & 0 & -0.62 \\ 0.38 & -0.78 & -0.49 \\ -0.48 & -0.62 & 0.62 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 1.62 & 0 \\ 0 & 1.0 \\ 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} -0.78 & 0.62 \\ -0.62 & -0.78 \end{bmatrix}}_{V^\top}. \tag{4.67b}$$

(Values are approximate.)

Example 4.13 (Computing the SVD)
Let us find the singular value decomposition of

$$A = \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}. \tag{4.81}$$

The SVD requires us to compute the right-singular vectors v_j, the singular values σ_k, and the left-singular vectors u_i.

Step 1: Right-singular vectors as the eigenbasis of $A^\top A$.
We start by computing

$$A^\top A = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & -2 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}. \tag{4.82}$$

We compute the singular values and right-singular vectors v_j through the eigenvalue decomposition of $A^\top A$, which is given as

$$A^\top A = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{30}} & \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{30}} \\ 0 & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} = PDP^\top, \tag{4.83}$$

and we obtain the right-singular vectors as the columns of P so that

$$V = P = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix}. \tag{4.84}$$

Step 2: Singular-value matrix.
As the singular values σ_i are the square roots of the eigenvalues of

1 Compute $A^\top A$ and its eigen-decomposition (right singular vectors)

$$A^\top A = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & -2 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \tag{4.82}$$

Now find eigenvalues and orthonormal eigenvectors of $A^\top A$.

- Eigenvalues: $\lambda_1 = 6,\ \lambda_2 = 1,\ \lambda_3 = 0.$
- One eigenvector for each:
  - $\lambda_1 = 6:\ v_1 \propto [5, -2, 1]^\top$
  - $\lambda_2 = 1:\ v_2 \propto [0, 1, 2]^\top$
  - $\lambda_3 = 0:\ v_3 \propto [-1, -2, 1]^\top$

Normalize them:

$$\|[5,-2,1]\| = \sqrt{30}, \quad \|[0,1,2]\| = \sqrt{5}, \quad \|[-1,-2,1]\| = \sqrt{6}$$

So an orthonormal eigenbasis is

$$P = V = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & -\frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix} \tag{4.84}$$

and

$$A^\top A = P \begin{bmatrix} 6 & 0 & 0 \\ 0 & 1 & 0 \\ \ldots \end{bmatrix} P^\top = PDP^\top \tag{4.83}$$

$$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \quad (4.83)$$

The columns of V are the **right singular vectors** v_1, v_2, v_3.

---

2 Singular values and Σ

Singular values are the square roots of the eigenvalues of $A^\top A$:

$$\sigma_1 = \sqrt{6}, \quad \sigma_2 = 1, \quad \sigma_3 = 0.$$

Since A is 2×3, we use the "thin" $\Sigma \in \mathbb{R}^{2\times 3}$:

$$\Sigma = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

---

2 Singular values and Σ

Singular values are the square roots of the eigenvalues of $A^\top A$:

$$\sigma_1 = \sqrt{6}, \quad \sigma_2 = 1, \quad \sigma_3 = 0.$$

Since A is 2×3, we use the "thin" $\Sigma \in \mathbb{R}^{2\times 3}$:

$$\Sigma = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

---

3 Left singular vectors u_i and U

For non-zero singular values,

$$u_i = \frac{1}{\sigma_i} A v_i.$$

For $\sigma_1 = \sqrt{6}$:

$$v_1 = \frac{1}{\sqrt{30}} \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \quad A v_1 = \frac{1}{\sqrt{30}} \begin{bmatrix} 6 \\ -12 \end{bmatrix} = \begin{bmatrix} 6/\sqrt{30} \\ -12/\sqrt{30} \end{bmatrix}.$$

Then

$$u_1 = \frac{1}{\sqrt{6}} A v_1 = \begin{bmatrix} 1/\sqrt{5} \\ -2/\sqrt{5} \end{bmatrix}.$$

For $\sigma_2 = 1$:

$$v_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \quad A v_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}.$$

So

$$u_2 = A v_2 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}.$$

These columns are orthonormal, so

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

(There is no third left singular vector because A is 2×3, rank 2.)

✅ Final SVD

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

with

$$U = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \cdots \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ \cdots \end{bmatrix}, \quad V = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & -\frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{6}} \\ \cdots \end{bmatrix}.$$

$$\begin{bmatrix} -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

You can check by multiplying $U\Sigma V^\top$ that you recover
$A = \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}$.

5.3:

Gradient of a Vector-Valued Linear Function

Let

$$f(x) = Ax, A = \begin{bmatrix} 2 & -1 & 0 \\ 3 & 1 & 4 \end{bmatrix}, x \in \mathbb{R}^3.$$

(a) Compute the Jacobian $\frac{df(x)}{dx}$.

(b) Verify it equals the matrix A.

Gradient of a Nonlinear Vector-Valued Function

Let

$$f(x) = \begin{bmatrix} x_1 x_2 + e^{x_3} \\ x_1^2 - 3x_2 + \sin\ x_3 \\ x_2 x_3^2 \end{bmatrix}.$$

Find the full Jacobian

$$J = \frac{df(x)}{dx} \in \mathbb{R}^{3\times 3}.$$

Jacobian Determinant (Area Scaling)

Given the linear mapping

$$y = Jx, J = \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix},$$

(a) Compute $\det(J)$.

(b) Interpret it geometrically as area scaling.

(c) Show how it transforms the unit square to a parallelogram.

Linear Regression Gradient Using Vector Calculus

Consider the linear model:

$$y = \Phi\theta,$$

where

where

- $\theta \in \mathbb{R}^2$ is the parameter vector,
- $\Phi \in \mathbb{R}^{3\times 2}$ is the design matrix,
- $y \in \mathbb{R}^3$ is the vector of observations.

Let:

$$\Phi = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 2 & -1 \end{bmatrix}, \qquad \theta = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \qquad y = \begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix}.$$

Define:

$$e(\theta) = y - \Phi\theta, \qquad L(\theta) = \|e(\theta)\|^2.$$

Task

Compute the gradient

$$\nabla_\theta L(\theta) = \frac{\partial L}{\partial \theta}$$

step by step.

✅ Step-by-Step Solution

1. Compute the prediction $\Phi\theta$

$$\Phi\theta = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1+6 \\ 3 \\ 2-3 \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \\ -1 \end{bmatrix}.$$

---

2. Compute the error vector $e(\theta) = y - \Phi\theta$

$$e(\theta) = \begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 7 \\ 3 \\ -1 \end{bmatrix} = \begin{bmatrix} -3 \\ -2 \\ 1 \end{bmatrix}.$$

Change of Coordinates (Basis Transformation)

Given bases

$$b_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, b_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

and new basis

$$c_1 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}, c_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

(a) Find the change-of-basis matrix J such that

$$Jb_1 = c_1, Jb_2 = c_2.$$

(b) Verify by computing the determinant that the area scaling equals 3.

Jacobian of Nonlinear Coordinate Transform

Given

$$y_1 = -2x_1 + x_2, y_2 = x_1 + x_2,$$

Compute the Jacobian

$$J = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{bmatrix},$$

and compute the determinant.

Chain Rule (Vector Form)

Let:

$$h(t) = (f \circ g)(t)$$

3. Formula for the gradient

From Example 5.11:

$$\frac{\partial L}{\partial \theta} = -2\Phi^T e(\theta).$$

$\partial\theta$

4. Compute Φ^T

$$\Phi^T = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & -1 \end{bmatrix}.$$

5. Multiply $\Phi^T e$

$$\Phi^T e = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & -1 \end{bmatrix} \begin{bmatrix} -3 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1(-3) + 0(-2) + 2(1) \\ 2(-3) + 1(-2) + (-1)(1) \end{bmatrix}.$$

Compute:

- First entry: $-3 + 2 = -1$
- Second entry: $-6 - 2 - 1 = -9$

Thus:

$$\Phi^T e = \begin{bmatrix} -1 \\ -9 \end{bmatrix}.$$

6. Compute the gradient

$$\frac{\partial L}{\partial \theta} = -2 \begin{bmatrix} -1 \\ -9 \end{bmatrix} = \begin{bmatrix} 2 \\ 18 \end{bmatrix}.$$

✅ Final Answer

$$\boxed{\nabla_\theta L(\theta) = \begin{bmatrix} 2 \\ 18 \end{bmatrix}}$$

This is the gradient of the least-squares loss with respect to the parameter vector θ.

Gradient of a Linear Vector Function

Given

$$f(x) = Ax, A = \begin{bmatrix} 2 & -1 & 0 \\ 3 & 1 & 4 \end{bmatrix}, x \in \mathbb{R}^3.$$

Write components:

$$f_1(x) = 2x_1 - x_2, f_2(x) = 3x_1 + x_2 + 4x_3.$$

The Jacobian is

$$\frac{df}{dx} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} \end{bmatrix}.$$

Compute:

- $\frac{\partial f_1}{\partial x_1} = 2, \frac{\partial f_1}{\partial x_2} = -1, \frac{\partial f_1}{\partial x_3} = 0$
- $\frac{\partial f_2}{\partial x_1} = 3, \frac{\partial f_2}{\partial x_2} = 1, \frac{\partial f_2}{\partial x_3} = 4$

Thus

$$\boxed{\frac{df}{dx} = \begin{bmatrix} 2 & -1 & 0 \\ 3 & 1 & 4 \end{bmatrix} = A}$$

This confirms: **the Jacobian of a linear function $f(x) = Ax$is exactly A.**

Jacobian of a Nonlinear Vector Function

$$f(x) = \begin{bmatrix} x_1x_2 + e^{x_3} \\ x_1^2 - 3x_2 + \sin\ x_3 \\ x_2x_3^2 \end{bmatrix}.$$

Compute all partial derivatives:

First component:

$$\frac{\partial f_1}{\partial x_1} = x_2, \frac{\partial f_1}{\partial x_2} = x_1, \frac{\partial f_1}{\partial x_3} = e^{x_3}.$$

Second:

$$\frac{\partial f_2}{\partial x_1} = 2x_1, \frac{\partial f_2}{\partial x_2} = -3, \frac{\partial f_2}{\partial x_3} = \cos\ x_3.$$

Third:

$$\frac{\partial f_3}{\partial x_1} = 0, \frac{\partial f_3}{\partial x_2} = x_3^2, \frac{\partial f_3}{\partial x_3} = 2x_2x_3.$$

Jacobian:

$$\boxed{J(x) = \begin{bmatrix} x_2 & x_1 & e^{x_3} \\ 2x_1 & -3 & \cos x_3 \\ 0 & x_3^2 & 2x_2x_3 \end{bmatrix}}$$

Determinant as Area Scaling

Given

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

Determinant:

$$\det(J) = (-2)(1) - (1)(1) = -3.$$

Geometric meaning:

$$|\det(J)| = 3$$

This means the linear map **scales all areas by a factor of 3** (and flips orientation because the determinant is negative).

Unit square → parallelogram:

Question

Let

1. $A \in \mathbb{R}^{3\times 2}$ and

2. Define functions

and

(a) Compute the gradient (Jacobian)

(b) Compute the gradient $\frac{dh}{dt}$ using t

Solutions

(a) Gradient of the vector-valued

Write A and x as

Then

$$f(x) =$$

The Jacobian (gradient) is the matrix

Compute each:

- $f_1(x) = A_{11}x_1 + A_{12}x_2$:
  $\dfrac{\partial f_1}{\partial x_1} = A_{11}, \ \dfrac{\partial f_1}{\partial x_2} = A_{12}.$
- $f_2(x) = A_{21}x_1 + A_{22}x_2$:
  $\dfrac{\partial f_2}{\partial x_1} = A_{21}, \ \dfrac{\partial f_2}{\partial x_2} = A_{22}.$
- $f_3(x) = A_{31}x_1 + A_{32}x_2$:
  $\dfrac{\partial f_3}{\partial x_1} = A_{31}, \ \dfrac{\partial f_3}{\partial x_2} = A_{32}.$

So

👉 **Result:** For $f(x) = Ax$, the grad

$$f(x) = Ax, \quad x \in \mathbb{R}^2,\ f(x) \in \mathbb{R}^3.$$

$$f : \mathbb{R}^2 \to \mathbb{R}, \qquad f(x_1, x_2) = x_1^2 e^{x_2},$$

$$g : \mathbb{R} \to \mathbb{R}^2, \qquad g(t) = \begin{bmatrix} t^3 \\ \sin t \end{bmatrix},$$

$$h(t) = (f \circ g)(t) = f(g(t)).$$

$\dfrac{\partial f}{\partial x}$ for $f(x) = Ax$.

the chain rule, writing all intermediate derivatives clearly.

---

d function $f(x) = Ax$

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

$$Ax = \begin{bmatrix} A_{11}x_1 + A_{12}x_2 \\ A_{21}x_1 + A_{22}x_2 \\ A_{31}x_1 + A_{32}x_2 \end{bmatrix} = \begin{bmatrix} f_1(x) \\ f_2(x) \\ f_3(x) \end{bmatrix}.$$

of partial derivatives:

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \\ \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} \end{bmatrix}.$$

$$\frac{\partial f}{\partial x} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix} = A.$$

lient/Jacobian is simply A.

(b) Chain rule for $h(t) = f(g(t))$

Recall

$$f(x_1, x_2) = x_1^2 e^{x_2}, \quad g(t)$$

So

$$h(t) = f($$

We'll compute $\frac{dh}{dt}$ using the **matrix chain rul**

Step 1: Gradient of f with respect to x

$$f$$

$$\frac{\partial f}{\partial x_1} = 2$$

So in vector form (row gradient, as in the text

$$\frac{\partial f}{\partial x} = [$$

Step 2: Gradient of g with respect to t

$$g(t) = \begin{bmatrix} t^3 \\ \sin \end{bmatrix}$$

Step 3: Chain rule

Matrix chain rule (like eq. 5.74a):

Plug in:

$$\frac{dh}{dt} = [$$

This is a 1×2 times 2×1 product → scalar:

$$\frac{dh}{dt} = 2x$$

Now substitute $x_1 = t^3,\ x_2 = \sin t$:

- $x_1 = t^3$
- $x_1^2 = t^6$
- $e^{x_2} = e^{\sin t}$

So

$$\frac{dh}{dt} = 2(t^3$$

$$= 6t$$

You can factor if you like:

$$\frac{dh}{dt} =$$

👉 **Final gradient:**

$$\boxed{\frac{dh}{dt} = 6t^5 e^{\sin t} +}$$

$$= \begin{bmatrix} t^3 \\ \sin t \end{bmatrix}, \quad x = g(t) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} t^3 \\ \sin t \end{bmatrix}.$$

$$g(t)) = (t^3)^2 e^{\sin t} = t^6 e^{\sin t}.$$

le, as in the book.

$$f(x_1, x_2) = x_1^2 e^{x_2}.$$

$$2x_1 e^{x_2}, \qquad \frac{\partial f}{\partial x_2} = x_1^2 e^{x_2}.$$

t:

$$\begin{bmatrix} 2x_1 e^{x_2} & x_1^2 e^{x_2} \end{bmatrix} \in \mathbb{R}^{1\times 2}.$$

$$t\end{bmatrix} \Rightarrow \frac{\partial g}{\partial t} = \begin{bmatrix} 3t^2 \\ \cos t \end{bmatrix} \in \mathbb{R}^{2\times 1}.$$

$$\frac{dh}{} = \frac{\partial f}{} \frac{\partial g}{}$$

$$\frac{}{dt} = \frac{}{\partial x}\frac{}{\partial t}.$$

$$2x_1 e^{x_2} \quad x_1^2 e^{x_2}\big] \begin{bmatrix} 3t^2 \\ \cos t \end{bmatrix}.$$

:

$$x_1 e^{x_2} \cdot 3t^2 + x_1^2 e^{x_2} \cdot \cos t.$$

$$^5) e^{\sin t} \cdot 3t^2 + (t^6) e^{\sin t} \cos t$$

$$t^5 e^{\sin t} + t^6 e^{\sin t} \cos t.$$

$$= t^5 e^{\sin t}(6 + t \cos t).$$

$$t^6 e^{\sin t} \cos t = t^5 e^{\sin t}(6 + t \cos t).$$

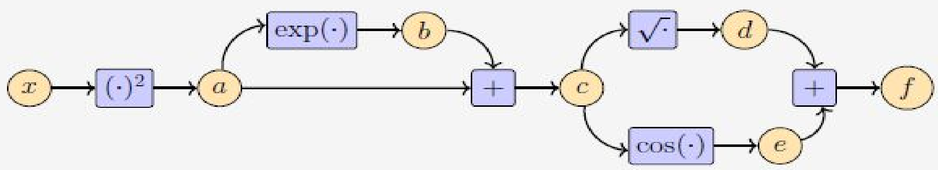
Example 5.14

Consider the function

$$f(x) = \sqrt{x^2 + \exp(x^2)} + \cos\left(x^2 + \exp(x^2)\right) \tag{5.122}$$

from (5.109). If we were to implement a function f on a computer, we would be able to save some computation by using *intermediate variables*:

$$a = x^2\,, \tag{5.123}$$
$$b = \exp(a)\,, \tag{5.124}$$
$$c = a + b\,, \tag{5.125}$$
$$d = \sqrt{c}\,, \tag{5.126}$$
$$e = \cos(c)\,, \tag{5.127}$$
$$f = d + e\,. \tag{5.128}$$



This is the same kind of thinking process that occurs when applying the chain rule. Note that the preceding set of equations requires fewer operations than a direct implementation of the function $f(x)$ as defined in (5.109). The corresponding computation graph in Figure 5.11 shows the flow of data and computations required to obtain the function value f.

The set of equations that include intermediate variables can be thought of as a computation graph, a representation that is widely used in implementations of neural network software libraries. We can directly compute the derivatives of the intermediate variables with respect to their corresponding inputs by recalling the definition of the derivative of elementary functions. We obtain the following:

$$\frac{\partial a}{\partial x} = 2x \tag{5.129}$$
$$\frac{\partial b}{\partial a} = \exp(a) \tag{5.130}$$

We have

$z =$

with

(A) Forward p

1. Compute z

2. Hidden activ

3. Output y

4. Loss L

$L =$

(B) Backpropa

We work backwar

1. $\partial L/\partial y$

$$= w_1 x + b_1, \quad h = \sigma(z) = \frac{1}{1+e^{-z}}, \quad y = w_2 h + b_2, \quad L = \tfrac{1}{2}(y-t)^2$$

$$x = 2,\ t = 1,\ w_1 = 0.5,\ b_1 = 0.1,\ w_2 = -1,\ b_2 = 0.2.$$

ɔass

$$z = w_1 x + b_1 = 0.5 \cdot 2 + 0.1 = 1.0 + 0.1 = 1.1$$

ation h

$$h = \sigma(z) = \frac{1}{1+e^{-1.1}} \approx 0.7503$$

$$y = w_2 h + b_2 = -1 \cdot 0.7503 + 0.2 \approx -0.5503$$

$$\tfrac{1}{2}(y-t)^2 = \tfrac{1}{2}(-0.5503-1)^2 = \tfrac{1}{2}(-1.5503)^2 \approx \tfrac{1}{2} \cdot 2.4033 \approx 1.2017$$

---

agation – gradients

rds using the chain rule.

$$L = \tfrac{1}{2}(y-t)^2 \quad \Rightarrow \quad \frac{\partial L}{\partial y} = y - t$$

$$\frac{\partial L}{\partial y} = -0.5503 - 1 \approx -1.5503$$

---

2. Gradients for w_2, b_2

$$y = w_2 h + b_2 \Rightarrow \frac{\partial y}{\partial w_2} = h, \quad \frac{\partial y}{\partial b_2} = 1, \quad \frac{\partial y}{\partial h} = w_2$$

So

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial w_2} = (y - t)\, h \approx (-1.5503)(0.7503) \approx -1.1$$

$$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial b_2} = (y - t) \cdot 1 \approx -1.5503$$

---

3. Gradient for h

$$\frac{\partial L}{\partial h} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial h} = (y - t)\, w_2 = (-1.5503)(-1) \approx 1.5503$$

---

4. Gradient for z

We use the sigmoid derivative:

$$h = \sigma(z), \quad \sigma'(z) = h(1 - h)$$

So

$$\frac{\partial h}{\partial z} = h(1 - h) \approx 0.7503(1 - 0.7503) \approx 0.7503 \cdot 0.2497 \approx 0$$

Then

$$\frac{\partial L}{\partial z} = \frac{\partial L}{\partial h}\frac{\partial h}{\partial z} \approx 1.5503 \cdot 0.1874 \approx 0.2905$$

631

5. Gradients for w_1, b_1

$$z = w_1 x + b_1 \Rightarrow \frac{\partial z}{\partial w_1} = x, \quad \frac{\partial z}{\partial b}$$

Therefore

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial w_1} = \frac{\partial L}{\partial z} x \approx 0.2905 \cdot 2$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial b_1} = \frac{\partial L}{\partial z} \cdot 1 \approx 0.2$$

(If you wanted, $\partial L/\partial x = \frac{\partial L}{\partial z} w_1 \approx 0.1452$.)

Summary of gradients (approx.)

$$\frac{\partial L}{\partial y} \approx -1.5503$$

$$\frac{\partial L}{\partial w_2} \approx -1.1631, \quad \frac{\partial L}{\partial b_2} \approx -1.5$$

$$\frac{\partial L}{\partial h} \approx 1.5503$$

$$\frac{\partial L}{\partial z} \approx 0.2905$$

$$\frac{\partial L}{\partial w_1} \approx 0.5809, \quad \frac{\partial L}{\partial b_1} \approx 0.2905$$

(C) Reverse-mode automatic differentiation (short exp

.1874

Reverse-mode automatic differentiation (AD) evaluates the computati values in a graph ("tape"), then runs a **backward pass** that applies the inputs. The backward pass reuses intermediates, so the cost of getting respect to many parameters is only a small constant factor more than Unlike symbolic differentiation, reverse-mode AD does not build huge numerical finite differences it is exact up to machine precision. That's v (backpropagation) is the standard method for training neural network

$\frac{z}{_1} = 1$

$2 \approx 0.5809$

2905

5503

5

---

lanation)

on **once forward**, storing intermediate chain rule from outputs back to all gradients of one scalar loss with the cost of computing the loss itself. algebraic expressions, and unlike why reverse-mode AD s with millions of parameters.

6.2.1

We observe two categorical random variables:

- **X** has 4 possible states:

$$x_1, x_2, x_3, x_4$$

- **Y** has 3 possible states:

$$y_1, y_2, y_3$$

The following table gives the **joint frequencies** n_{ij}:

Using this table, answer ALL of the following:

X/Y	y_1	y_2	y_3	Row sum (r_i)
x_1	4	5	1	?
x_2	2	6	2	?
x_3	3	1	4	?
x_4	1	3	2	?
Column sum	?	?	?	N = ?

(A) Compute all row sums r_i.

(B) Compute all column sums c_j.

(C) Compute the total number of observations N.

(D) Compute the joint probabilities:

$$P(X = x_2, Y = y_2), P(X = x_4, Y = y_3)$$

(F) Conditional probabilities

Formula:

$$P(Y = y_j \mid X = x_i) = \frac{n_{ij}}{r_i}$$
$$P(X = x_i \mid Y = y_j) = \frac{n_{ij}}{c_j}$$

$P(Y = y_2 \mid X = x_1) = n_{12}/r_1 = 5/10 = 0.5$

$P(X = x_2 \mid Y = y_3) = n_{23}/c_3 = 2/9 \approx 0.222$

(G) Check summation rules

$$\sum_i P(X = x_i)$$

$$10/34 + 10/34 + 8/34 + 6/34 = 34/34 = 1$$

Correct ✓

$$\sum_j P(Y = y_j)$$

$$10/34 + 15/34 + 9/34 = 34/34 = 1$$

Correct ✓

(H) Interpretation

✔ Most frequent pair (largest n_{ij})

- Largest cell = 6 (for x_2, y_2)

So the pair ($X=x_2$, $Y=y_2$) is the most common.

Compute the marginal probabilities:

$$P(X = x_3), P(Y = y_1)$$

Compute the conditional probabilities:

1. $P(Y = y_2 \mid X = x_1)$
2. $P(X = x_2 \mid Y = y_3)$

Check that

$$\sum_i P(X = x_i) = 1, \sum_j P(Y = y_j) = 1$$

Interpret the result:

ch pair (X and Y states) seems the most frequent?
ch category seems most probable for X? For Y?

Row sums

$\cdot + 5 + 1 = 10$

$\cdot + 6 + 2 = 10$

$\cdot + 1 + 4 = 8$

$. + 3 + 2 = 6$

Column sums

$\cdot + 2 + 3 + 1 = 10$

$\cdot + 6 + 1 + 3 = 15$

$. + 2 + 4 + 2 = 9$

(C) Total

(or sum of columns: 10·

(D) Joint probabilities

Formula:

$P(X = x_2, Y = y_2) =$

$P(X = x_4, Y = y_3) =$

(E) Marginal probabil

Formula:

$P(X = x_3) = r_3/N =$

$P(Y = y_1) = c_1/N =$

---

st probable category of X:

$$\max\{10/34, 10/34, 8/34, 6/34\} = 10/34$$

x_1 and x_2 are tied as most probable.

---

st probable category of Y:

$$\max\{10/34,\ 15/34,\ 9/34\} = 15/34$$

y_2 is the most probable Y category

FORMULAS :

1. Joint Probability (Pr

Where:

- n_{ij}= frequency o
- N= total observa

---

2. Marginal Probabilit

Where:

(column sum)

3. Marginal Probability

Where:

(row sum)

4. Probability Mass Fu

$$N = 10 + 10 + 8 + 6 = 34$$

+15+9 = 34)

:

$$P(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

$n_{22}/34 = 6/34 \approx 0.176$

$n_{43}/34 = 2/34 \approx 0.0588$

---

lities

$$P(X = x_i) = \frac{r_i}{N}$$
$$P(Y = y_j) = \frac{c_j}{N}$$

$8/34 \approx 0.235$

$10/34 \approx 0.294$

X \ Y	y_1	y_2
x_1	4	2
x_2	5	1
x_3	3	6
x_4	2	3
x_5	1	4
Column Sum c_j	15	16

So total number of events:

---

1. Joint Probability

Joint probability:

$p($

Example:

$p($

---

2. Marginal Probability of X

$P(X$

Compute each:

- $P(X = x_1) = \frac{9}{45} = 0.20$
- $P(X = x_2) = \frac{10}{45} = 0.222$
- $P(X = x_3) = \frac{10}{45} = 0.222$
- $P(X = x_4) = \frac{8}{45} = 0.178$
- $P(X = x_5) = \frac{8}{45} = 0.178$

Check sum:

$$0.20 + 0.222 + 0$$

---

3. Marginal Probability of Y

$P(Y$

Compute each:

- $P(Y = y_1) = \frac{15}{45} = 0.333$
- $P(Y = y_2) = \frac{16}{45} = 0.355$
- $P(Y = y_3) = \frac{14}{45} = 0.311$

Sum:

$$0.333 + 0$$

---

robability of both X and Y together)

$$P(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

of (X = x_i AND Y = y_j)

tions

---

ty of X

$$P(X = x_i) = \frac{c_i}{N}$$

$$c_i = \sum^{m} n_{ij}$$

$j=1$

---

y of Y

$$P(Y=y_j)=\frac{r_j}{N}$$

$$r_j=\sum_{i=1}^{n} n_{ij}$$

ınction (PMF)

4. Conditional Probability

Example:

Probability of $Y=y_2$ given $X=x_3$:

$$P(Y=y_2 \mid X=$$

Another example:

$$P(Y=y_1 \mid$$

---

y_3	Row Sum r_i
3	9
4	10
1	10
3	8
3	8
14	45 (Total N)

$N = 45$

$x_i, y_j) = \frac{n_{ij}}{N}$

$x_1, y_2) = \frac{2}{45}$

$X = x_i) = \frac{r_i}{N}$

$0.222 + 0.178 + 0.178 = 1$

$Y = y_j) = \frac{c_j}{N}$

$0.355 + 0.311 = 1$

$$= x_3) = \frac{n_{32}}{r_3} = \frac{6}{10} = 0.6$$

$$X = x_2) = \frac{5}{10} = 0.5$$

Example 7.1
Consider a quadratic function in two dimensions

$$f\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix}x_1\\x_2\end{bmatrix}^\top \begin{bmatrix}2 & 1\\1 & 20\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} - \begin{bmatrix}5\\3\end{bmatrix}^\top \begin{bmatrix}x_1\\x_2\end{bmatrix} \tag{7.7}$$

with gradient

$$\nabla f\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \begin{bmatrix}x_1\\x_2\end{bmatrix}^\top \begin{bmatrix}2 & 1\\1 & 20\end{bmatrix} - \begin{bmatrix}5\\3\end{bmatrix}^\top . \tag{7.8}$$

Starting at the initial location $x_0 = [-3, -1]^\top$, we iteratively apply (7.6) to obtain a sequence of estimates that converge to the minimum value

Let's calculate the first iteration of gradient descent using the numbers from your image.

1. Identify the parameters: From Equation (7.7), we have the matrix A and vector $\mathbf{b}$:

$$A = \begin{bmatrix}2 & 1\\1 & 20\end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix}5\\3\end{bmatrix}$$

2. Identify the Start Point: The text states the initial location is:

$$\mathbf{x}_0 = \begin{bmatrix}-3\\-1\end{bmatrix}$$

3. Calculate the Gradient at $\mathbf{x}_0$: Using Equation (7.8) from the image (written in standard column vector form):

$$\nabla f(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$$

First, multiply A by $\mathbf{x}_0$:

$$A\mathbf{x}_0 = \begin{bmatrix}2 & 1\\1 & 20\end{bmatrix}\begin{bmatrix}-3\\-1\end{bmatrix} = \begin{bmatrix}(2)(-3) + (1)(-1)\\(1)(-3) + (20)(-1)\end{bmatrix} = \begin{bmatrix}-7\\-23\end{bmatrix}$$

Next, subtract $\mathbf{b}$:

$$\nabla f(\mathbf{x}_0) = \begin{bmatrix}-7\\-23\end{bmatrix} - \begin{bmatrix}5\\3\end{bmatrix} = \begin{bmatrix}-12\\-26\end{bmatrix}$$

The gradient is $[-12, -26]^T$. This tells us the slope is very steep in the negative direction.

4. Update the Position (The Step): We need a learning rate (α). The image doesn't provide one, so let's assume a standard rate of $\alpha = 0.05$.

$$\mathbf{x}_1 = \mathbf{x}_0 - 0.05 \times \nabla f(\mathbf{x}_0)$$

$$\mathbf{x}_1 = \begin{bmatrix}-3\\-1\end{bmatrix} - 0.05\begin{bmatrix}-12\\-26\end{bmatrix}$$

$$\begin{bmatrix}-3\end{bmatrix} \quad \begin{bmatrix}-0.6\end{bmatrix}$$

$$\mathbf{x}_1 = \begin{bmatrix} 3 \\ -1 \end{bmatrix} - \begin{bmatrix} 0.6 \\ -1.3 \end{bmatrix}$$

$$\mathbf{x}_1 = \begin{bmatrix} -3 + 0.6 \\ -1 + 1.3 \end{bmatrix} = \begin{bmatrix} -2.4 \\ 0.3 \end{bmatrix}$$

1. Define the quadratic function

We have

$$f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 20 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 5 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Let's expand it so it looks more familiar.

1. Multiply the matrix by the vector:

$$\begin{bmatrix} 2 & 1 \\ 1 & 20 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + x_2 \\ x_1 + 20x_2 \end{bmatrix}$$

2. Left-multiply by $[x_1\ x_2]$:

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 2x_1 + x_2 \\ x_1 + 20x_2 \end{bmatrix} = x_1(2x_1 + x_2) + x_2(x_1 + 20x_2) = 2x_1^2 + 2x_1x_2 + 20x_2^2$$

3. Multiply by $\frac{1}{2}$:

$$\frac{1}{2}(\dots) = x_1^2 + x_1x_2 + 10x_2^2$$

4. Last term:

$$\begin{bmatrix} 5 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 5x_1 + 3x_2$$

So

$$\boxed{f(x_1, x_2) = x_1^2 + x_1x_2 + 10x_2^2 - 5x_1 - 3x_2}$$

---

2. Gradient of f

Take partial derivatives:

$$\frac{\partial f}{\partial x_1} = 2x_1 + x_2 - 5$$
$$\frac{\partial f}{\partial x_2} = x_1 + 20x_2 - 3$$

So

$$\boxed{\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 + x_2 - 5 \\ x_1 + 20x_2 - 3 \end{bmatrix}}$$

This matches the compact matrix form in your book:

$$\nabla f(x) = \begin{bmatrix} 2 & 1 \\ 1 & 20 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

---

3. Gradient Descent Example

Use gradient descent with:

- learning rate $\alpha = 0.05$
- starting point $x^{(0)} = \begin{bmatrix} -3 \\ -1 \end{bmatrix}$

Update rule:

$$x^{(k+1)} = x^{(k)} - \alpha \nabla f\left(x^{(k)}\right)$$

Step 0 → 1

Start: $x^{(0)} = (-3, -1)$

Gradient at $x^{(0)}$:

$$\nabla f(-3,-1) = \begin{bmatrix} 2(-3) + (-1) - 5 \\ (-3) + 20(-1) - 3 \end{bmatrix} = \begin{bmatrix} -12 \\ -26 \end{bmatrix}$$

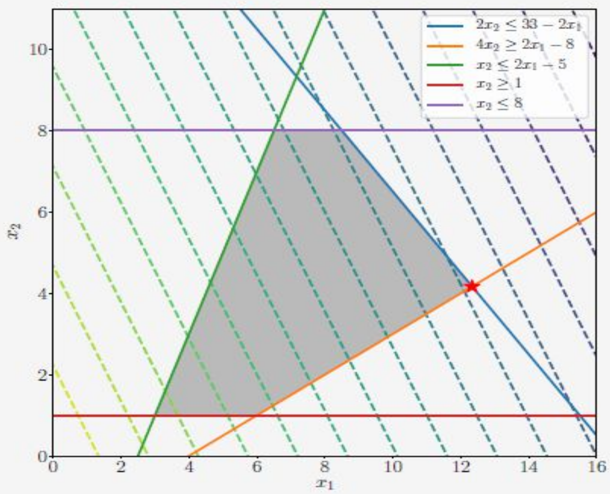
Update:

$$x^{(1)} = x^{(0)} - 0.05 \cdot (-12, -26) = (-3, -1) + (0.6, 1.3) = \boxed{(-2.4,\ 0.3)}$$

Example 7.5 (Linear Program)
Consider the linear program

$$
\begin{aligned}
\min_{\boldsymbol{x}\in\mathbb{R}^2} \quad & -\begin{bmatrix}5\\3\end{bmatrix}^\top \begin{bmatrix}x_1\\x_2\end{bmatrix} \\
\text{subject to} \quad & \begin{bmatrix}2 & 2\\ 2 & -4\\ -2 & 1\\ 0 & -1\\ 0 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} \leqslant \begin{bmatrix}33\\8\\5\\-1\\8\end{bmatrix}
\end{aligned}
\tag{7.44}
$$

with two variables. This program is also shown in Figure 7.9. The objective function is linear, resulting in linear contour lines. The constraint set in standard form is translated into the legend. The optimal value must lie in the shaded (feasible) region, and is indicated by the star.



3 Evaluate the objective at each vertex

Recall $f(x_1, x_2) = -5x_1 - 3x_2$.

1. At $\left(\frac{37}{3}, \frac{25}{6}\right)$:

$$f = -5 \cdot \frac{37}{3} - 3 \cdot \frac{25}{6} = -\frac{185}{3} - \frac{75}{6} = -\frac{445}{6} \approx -74.17$$

2. At $(17/2, 8)$:
   $f = -5 \cdot \frac{17}{2} - 3 \cdot 8 = -\frac{85}{2} - 24 = -66.5$
3. At $(6, 1)$:
   $f = -5 \cdot 6 - 3 \cdot 1 = -33$
4. At $(-2, 1)$:
   $f = -5(-2) - 3 \cdot 1 = 10 - 3 = 7$
5. At $(3/2, 8)$:
   $f = -5 \cdot \frac{3}{2} - 3 \cdot 8 = -\frac{15}{2} - 24 = -31.5$

We are minimizing, so we choose the smallest value:

$$\boxed{f_{\min} = -\frac{445}{6} \approx -74.17}$$

at

$$\boxed{x^* = \left(\frac{37}{3}, \frac{25}{6}\right) \approx (12.33,\ 4.17)}.$$

That point is the star shown in the plot: it lies in the feasible region and is where the objective's cont is as "low" (in the minimize sense) as possible while still touching the shaded region.

1 Write the linear program in inequalities

From the matrix form (7.44):

$$\min_{x\in\mathbb{R}^2} -[5\ 3]^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{s.t.} \quad \begin{bmatrix} 2 & 2 \\ 2 & -4 \\ -2 & 1 \\ 0 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq \begin{bmatrix} 33 \\ 8 \\ 5 \\ -1 \\ 8 \end{bmatrix}$$

Write constraints explicitly:

1. $2x_1 + 2x_2 \leq 33$
2. $2x_1 - 4x_2 \leq 8$
3. $-2x_1 + x_2 \leq 5$
4. $-x_2 \leq -1 \ \Rightarrow \ x_2 \geq 1$
5. $x_2 \leq 8$

Objective:

$$\min\ f(x_1, x_2) = -5x_1 - 3x_2$$

(Equivalently, maximize $5x_1 + 3x_2$.)

The shaded polygon in the figure is the set of points satisfying all 5 inequalities (the feasible region).

---

2 Find the corner points (vertices) of the feasible region

In 2-D linear programming, the optimum (if it exists and the region is bounded in the direction of optimization) occurs at a **vertex** of the feasible polygon.

Vertices are intersections of pairs of boundary lines:

- (1) $2x_1 + 2x_2 = 33$
- (2) $2x_1 - 4x_2 = 8$
- (3) $-2x_1 + x_2 = 5$
- (4) $x_2 = 1$
- (5) $x_2 = 8$

Solve intersections and keep only those that satisfy **all** inequalities.

The feasible vertices are:

1. Intersection of (1) & (2):

$$\begin{cases} 2x_1 + 2x_2 = 33 \\ 2x_1 - 4x_2 = 8 \end{cases} \Rightarrow x_1 = \frac{37}{3}, \quad x_2 = \frac{25}{6}$$

2. (1) & (5): $(x_1, x_2) = (17/2,\ 8)$
3. (2) & (4): $(x_1, x_2) = (6,\ 1)$
4. (3) & (4): $(x_1, x_2) = (-2,\ 1)$
5. (3) & (5): $(x_1, x_2) = (3/2,\ 8)$

All five lie in the shaded region.

[redacted]

our line

3 Evaluate the objective at each vertex

Recall $f(x_1, x_2) = -5x_1 - 3x_2$.

1. At $\left(\frac{37}{3}, \frac{25}{6}\right)$:

$$f = -5 \cdot \frac{37}{3} - 3 \cdot \frac{25}{6} = -\frac{185}{3} - \frac{75}{6} = -\frac{445}{6} \approx -74.17$$

2. At $(17/2, 8)$:
   $f = -5 \cdot \frac{17}{2} - 3 \cdot 8 = -\frac{85}{2} - 24 = -66.5$
3. At $(6, 1)$:
   $f = -5 \cdot 6 - 3 \cdot 1 = -33$
4. At $(-2, 1)$:
   $f = -5(-2) - 3 \cdot 1 = 10 - 3 = 7$
5. At $(3/2, 8)$:
   $f = -5 \cdot \frac{3}{2} - 3 \cdot 8 = -\frac{15}{2} - 24 = -31.5$

We are minimizing, so we choose the smallest value:

$$\boxed{f_{\min} = -\frac{445}{6} \approx -74.17}$$

at

$$\boxed{x^* = \left(\frac{37}{3}, \frac{25}{6}\right) \approx (12.33,\ 4.17)}.$$

That point is the star shown in the plot: it lies in the feasible region and is where the objective's contour li
is as "low" (in the minimize sense) as possible while still touching the shaded region.

[redacted]

ine