Student's Solutions Manual

Linear Algebra, Data Science, and Machine Learning

Springer, 2025

Jeff Calder Peter J. Olver

December 4, 2025

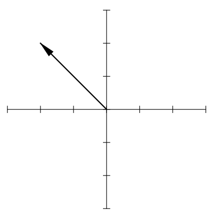
Contents

Chapter 1

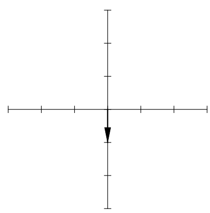
Vectors

1.1. Plot the following vectors in $\mathbb{R}^2$.
(a) ♡ $(-2,2)^T$, (b) ♡ $(0,-1)^T$, (c) ♢ $3\,(1,1)^T$, (d) $(-2,3)^T - (-5,3)^T$.

Solution: (a)



(b)



1.2. Suppose $\mathbf{v} = (1,2,-1)^T$ and $\mathbf{w} = (0,-1,2)^T$. Determine the following vectors:
(a) ♡ $-\mathbf{v}$, (b) $3\mathbf{v}$, (c) ♡ $-5\,\mathbf{w}$, (d) ♢ $\mathbf{v}+\mathbf{w}$, (e) $\mathbf{v}-\mathbf{w}$, (f) $2\mathbf{v}-3\mathbf{w}$.

Solution: (a) $(-1,-2,1)^T$, (c) $(0,5,-10)^T$.

1.3. Prove the arithmetic properties (a) ♡, (b) ♢, (c) ♡, (d), (e), (f) ♡, (g) for vectors in $\mathbb{R}^n$.

Solution: Given $\mathbf{u} = (u_1,u_2,\ldots,u_n)^T$, $\mathbf{v} = (v_1,v_2,\ldots,v_n)^T$, $\mathbf{w} = (w_1,w_2,\ldots,w_n)^T$, the i-th entry of each side of the equation are equal by the laws of arithmetic of real numbers: (a) $v_i+w_i = w_i+v_i$; (c) $(c+d)\,v_i = c\,v_i + d\,v_i$ and $c\,(v_i+w_i) = c\,v_i + c\,w_i$; (f) $v_i+0 = v_i = 0+v_i$, $v_i + (-v_i) = 0 = (-v_i)+v_i$. If $u_i = v_i + w_i$ then $w_i = v_i - u_i = v_i + (-u_i)$.

2.1. ♡ (a) Prove that the set of all vectors $(x,y,z)^T$ such that $x-y+4z=0$ forms a subspace of $\mathbb{R}^3$. (b) Explain why the set of all vectors that satisfy $x-y+4z=1$ does not form a subspace.

Solution: (a) If $\mathbf{v} = (x,y,z)^T$ satisfies $x-y+4z=0$ and $\widetilde{\mathbf{v}} = (\widetilde{x},\widetilde{y},\widetilde{z})^T$ also satisfies $\widetilde{x}-\widetilde{y}+4\widetilde{z}=0$, so does $\mathbf{v}+\widetilde{\mathbf{v}} = (x+\widetilde{x}, y+\widetilde{y}, z+\widetilde{z})^T$ since

$$(x+\widetilde{x}) - (y+\widetilde{y}) + 4\,(z+\widetilde{z}) = (x-y+4z) + (\widetilde{x}-\widetilde{y}+4\widetilde{z}) = 0,$$

as does $c\,\mathbf{v} = (cx,cy,cz)^T$ since

$$(cx) - (cy) + 4\,(cz) = c\,(x-y+4z) = 0. \qquad \blacksquare$$

(b) For instance, the zero vector $\mathbf{0} = (0,0,0)^T$ does not satisfy the equation.

2.2. Which of the following are subspaces of $\mathbb{R}^3$? Justify your answers! (a) ♡ The set of all vectors $(x, y, z)^T$ satisfying $x+y+z+1=0$. (b) ♢ The set of vectors of the form $(t, -t, 0)^T$ for $t \in \mathbb{R}$. (c) ♡ The set of vectors of the form $(r-s, r+2s, -s)^T$ for $r, s \in \mathbb{R}$. (d) The set of vectors whose first component equals 0. (e) The set of vectors whose last component equals 1. (f) ♡ The set of all vectors $(x, y, z)^T$ with $x \geq y \geq z$. (g) ♡ The set of all solutions to the equation $z = x - y$. (h) ♢ The set of all solutions to the equation $z = xy$. (i) The set of all solutions to the equation $x^2 + y^2 + z^2 = 0$. (j) The set of all solutions to the system $xy = yz = xz$.

Solution: (a) Not a subspace; (c) subspace; (f) not a subspace; (g) subspace.

2.3. Determine which of the following sets of vectors $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ are subspaces of $\mathbb{R}^n$: (a) ♡ all equal entries $x_1 = \cdots = x_n$; (b) ♡ all positive entries: $x_i \geq 0$; (c) ♢ first and last entries equal to zero: $x_1 = x_n = 0$; (d) the entries add up to zero: $x_1 + \cdots + x_n = 0$; (e) first and last entries differ by one: $x_1 - x_n = 1$.

Solution: (a) Subspace; (b) not a subspace.

2.6. Show that if V and W are subspaces of $\mathbb{R}^n$, then (a) ♡ their *intersection* $V \cap W$ is a subspace; (b) their *sum* $V + W = \{\mathbf{v} + \mathbf{w} \mid \mathbf{v} \in V,\ \mathbf{w} \in W\}$ is a subspace; but (c) ♢ their *union* $V \cup W$ is not a subspace, unless $V \subset W$ or $W \subset V$.

Solution: (a) If $\mathbf{v}, \mathbf{w} \in W \cap Z$, then $\mathbf{v}, \mathbf{w} \in W$, so $c\mathbf{v} + d\mathbf{w} \in W$ because W is a subspace, and $\mathbf{v}, \mathbf{w} \in Z$, so $c\mathbf{v} + d\mathbf{w} \in Z$ because Z is a subspace, hence $c\mathbf{v} + d\mathbf{w} \in W \cap Z$. ■

3.1. ♡ Show that $\begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix}$ belongs to the subspace of $\mathbb{R}^3$ spanned by $\begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 \\ -4 \\ 1 \end{pmatrix}$ by writing it as a linear combination of the spanning vectors.

Solution: $\begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix} = 2\begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix} - \begin{pmatrix} 5 \\ -4 \\ 1 \end{pmatrix}$.

3.3. Which of the following sets of vectors span all of $\mathbb{R}^2$? (a) ♡ $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$; (b) ♡ $\begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \end{pmatrix}$; (c) ♢ $\begin{pmatrix} 6 \\ -9 \end{pmatrix}, \begin{pmatrix} -4 \\ 6 \end{pmatrix}$; (d) $\begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \end{pmatrix}$; (e) ♡ $\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 4 \\ 8 \end{pmatrix}$; (f) $\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix}$.

Solution: (a) No, (b) yes, (e) no.

3.4. Determine whether the given vectors are linearly independent or linearly dependent:

$(a)\heartsuit \begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}2\\1\end{pmatrix}$, $(b)\heartsuit \begin{pmatrix}1\\3\end{pmatrix}, \begin{pmatrix}-2\\-6\end{pmatrix}$, $(c) \begin{pmatrix}2\\1\end{pmatrix}, \begin{pmatrix}-1\\3\end{pmatrix}, \begin{pmatrix}5\\2\end{pmatrix}$, $(d)\heartsuit \begin{pmatrix}1\\3\\-2\end{pmatrix}, \begin{pmatrix}0\\2\\-1\end{pmatrix}$,

$(e)\diamondsuit \begin{pmatrix}0\\1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\2\end{pmatrix}$, $(f)\diamondsuit \begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}0\\1\\1\end{pmatrix}$, $(g) \begin{pmatrix}4\\2\\0\\-6\end{pmatrix}, \begin{pmatrix}-6\\-3\\0\\9\end{pmatrix}$.

Solution: (a) Linearly independent; (b) linearly dependent; (d) linearly independent.

3.7. $\heartsuit$ Prove or give a counterexample to the following statement: If $\mathbf{v}_1, \ldots, \mathbf{v}_k$ do not span $\mathbb{R}^n$, then $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent.

Solution: False. For example, any set containing the zero vector that does not span $\mathbb{R}^n$ is linearly dependent. ■

4.1. Determine which of the following sets of vectors are bases of $\mathbb{R}^2$: $(a)\heartsuit \begin{pmatrix}2\\1\end{pmatrix}$;

$(b)\heartsuit \begin{pmatrix}1\\-1\end{pmatrix}, \begin{pmatrix}-1\\1\end{pmatrix}$; $(c)\diamondsuit \begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}2\\1\end{pmatrix}$; $(d)\heartsuit \begin{pmatrix}3\\5\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}$; $(e) \begin{pmatrix}2\\0\end{pmatrix}, \begin{pmatrix}-1\\2\end{pmatrix}, \begin{pmatrix}0\\-1\end{pmatrix}$.

Solution: (a) Not a basis; (b) basis; (d) not a basis.

4.2. Determine which of the following are bases of $\mathbb{R}^3$: $(a)\heartsuit \begin{pmatrix}2\\1\\5\end{pmatrix}, \begin{pmatrix}1\\5\\2\end{pmatrix}$; $(b)\heartsuit \begin{pmatrix}0\\1\\-5\end{pmatrix}$,

$\begin{pmatrix}-1\\3\\0\end{pmatrix}, \begin{pmatrix}1\\3\\0\end{pmatrix}$; $(c)\diamondsuit \begin{pmatrix}-1\\0\\1\end{pmatrix}, \begin{pmatrix}0\\4\\-1\end{pmatrix}, \begin{pmatrix}1\\-4\\0\end{pmatrix}$; $(d) \begin{pmatrix}2\\0\\-2\end{pmatrix}, \begin{pmatrix}-1\\2\\-1\end{pmatrix}, \begin{pmatrix}0\\-1\\0\end{pmatrix}, \begin{pmatrix}-1\\2\\1\end{pmatrix}$.

Solution: (a) Not a basis; (b) basis.

4.4. Find a basis for the following planes in $\mathbb{R}^3$:

$(a)\heartsuit$ the xy plane; (b) $z - 2y = 0$; $(c)\diamondsuit$ $4x + 3y - z = 0$.

Solution: $(a) \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}$.

4.5. $\heartsuit$ Show, by computing an example, how the uniqueness result in Proposition 1.18 fails if one has a linearly dependent set of vectors.

Solution: For instance, if

$$\mathbf{v}_1 = \begin{pmatrix}1\\0\end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix}0\\1\end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix}1\\1\end{pmatrix}, \qquad \text{then} \qquad \begin{pmatrix}2\\1\end{pmatrix} = 2\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_1 + \mathbf{v}_3.$$

In fact, there are infinitely many different ways of writing this vector as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. ■

Chapter 2

Inner Product, Orthogonality, Norm

1.1. Which of the following formulas for $\langle \mathbf{v}, \mathbf{w} \rangle$ define inner products on $\mathbb{R}^2$?
(a) ♡ $2v_1w_1 + 3v_2w_2$, (b) ♡ $v_1w_2 + v_2w_1$, (c) $(v_1+v_2)(w_1+w_2)$, (d) $v_1^2w_1^2 + v_2^2w_2^2$,
(e) ♢ $2v_1w_1 + (v_1-v_2)(w_1-w_2)$, (f) $4v_1w_1 - 2v_1w_2 - 2v_2w_1 + 4v_2w_2$.

Solution: (a) Yes; (b) no — not positive definite.

1.3. Prove that each of the following formulas for $\langle \mathbf{v}, \mathbf{w} \rangle$ defines an inner product on $\mathbb{R}^3$. Verify all the inner product axioms in careful detail:
(a) ♡ $v_1w_1 + 2v_2w_2 + 3v_3w_3$, (b) $4v_1w_1 + 2v_1w_2 + 2v_2w_1 + 4v_2w_2 + v_3w_3$,
(c) ♢ $2v_1w_1 - 2v_1w_2 - 2v_2w_1 + 3v_2w_2 - v_2w_3 - v_3w_2 + 2v_3w_3$.

Solution: (a) Bilinearity:

$$\begin{aligned}
\langle c\mathbf{u} + d\mathbf{v}, \mathbf{w} \rangle &= (cu_1 + dv_1)w_1 + 2(cu_2 + dv_2)w_2 + 3(cu_3 + dv_3)w_3 \\
&= c(u_1w_1 + 2u_2w_2 + 3u_3w_3) + d(v_1w_1 + 2v_2w_2 + 3v_3w_3) \\
&= c\langle \mathbf{u}, \mathbf{w} \rangle + d\langle \mathbf{v}, \mathbf{w} \rangle, \\
\langle \mathbf{u}, c\mathbf{v} + d\mathbf{w} \rangle &= u_1(cv_1 + dw_1) + 2u_2(cv_2 + dw_2) + 3u_3(cv_3 + dw_3) \\
&= c(u_1v_1 + 2u_2v_2 + 3u_3v_3) + d(u_1w_1 + 2u_2w_2 + 3u_3w_3) \\
&= c\langle \mathbf{u}, \mathbf{v} \rangle + d\langle \mathbf{u}, \mathbf{w} \rangle.
\end{aligned}$$

Symmetry: $\langle \mathbf{v}, \mathbf{w} \rangle = v_1w_1 + 2v_2w_2 + 3v_3w_3 = w_1v_1 + 2w_2v_2 + 3w_3v_3 = \langle \mathbf{w}, \mathbf{v} \rangle$.
Positivity: $\langle \mathbf{v}, \mathbf{v} \rangle = v_1^2 + 2v_2^2 + 3v_3^2 > 0$ for all $\mathbf{v} = (v_1, v_2, v_3)^T \neq \mathbf{0}$, because it is a sum of nonnegative terms, at least one of which is strictly positive. ■

1.4. Prove that the following quadratic forms on $\mathbb{R}^3$ are positive definite by writing each as a sum of squares. Then write down the corresponding inner product.
(a) ♡ $x^2 + 4xz + 3y^2 + 5z^2$, (b) ♢ $x^2 + 3xy + 3y^2 - 2xz + 8z^2$,
(c) $2x_1^2 + x_1x_2 - 2x_1x_3 + 2x_2^2 - 2x_2x_3 + 2x_3^2$.

Solution: (a) $(x+2z)^2 + 3y^2 + z^2$, $\langle \mathbf{v}, \mathbf{w} \rangle = v_1w_1 + 2v_1w_2 + 2v_2w_1 + 3v_2w_2 + 5v_3w_3$.

1.6. (a) ♡ Prove that $\langle \mathbf{x}, \mathbf{v} \rangle = 0$ for all $\mathbf{v} \in \mathbb{R}^n$ if and only if $\mathbf{x} = \mathbf{0}$. (b) ♢ Prove that $\langle \mathbf{x}, \mathbf{v} \rangle = \langle \mathbf{y}, \mathbf{v} \rangle$ for all $\mathbf{v} \in \mathbb{R}^n$ if and only if $\mathbf{x} = \mathbf{y}$. (c) Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a basis for $\mathbb{R}^n$. Prove that $\langle \mathbf{x}, \mathbf{v}_i \rangle = \langle \mathbf{y}, \mathbf{v}_i \rangle$ for all $i = 1, \ldots, n$ if and only if $\mathbf{x} = \mathbf{y}$.

Solution: (a) Choosing $\mathbf{v} = \mathbf{x}$, we have $0 = \langle \mathbf{x}, \mathbf{x} \rangle = \| \mathbf{x} \|^2$, and hence $\mathbf{x} = \mathbf{0}$. ■

1.7. Let $\langle \cdot, \cdot \rangle$ be an inner product on $\mathbb{R}^n$ and let $\| \cdot \|$ be the induced norm.

(a) ♡ Show that the norm satisfies the *parallelogram identity*

$$\| \mathbf{v} + \mathbf{w} \|^2 + \| \mathbf{v} - \mathbf{w} \|^2 = 2 \| \mathbf{v} \|^2 + 2 \| \mathbf{w} \|^2 \quad \text{for all} \quad \mathbf{v}, \mathbf{w} \in \mathbb{R}^n. \tag{2.24}$$

(b) ♢ Prove the identity

$$\langle \mathbf{v}, \mathbf{w} \rangle = \tfrac{1}{4} \left(\| \mathbf{v} + \mathbf{w} \|^2 - \| \mathbf{v} - \mathbf{w} \|^2 \right), \tag{2.25}$$

which allows one to reconstruct an inner product from its norm.

(c) Use (2.25) to find the inner product on $\mathbb{R}^2$ corresponding to the norm

$$\| \mathbf{v} \| = \sqrt{v_1^2 - 3 v_1 v_2 + 5 v_2^2}\,.$$

Solution: (a) $\| \mathbf{v} + \mathbf{w} \|^2 + \| \mathbf{v} - \mathbf{w} \|^2 = \langle \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w} \rangle + \langle \mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w} \rangle$
$= \left(\langle \mathbf{v}, \mathbf{v} \rangle + 2 \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle \right) + \left(\langle \mathbf{v}, \mathbf{v} \rangle - 2 \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle \right) = 2 \| \mathbf{v} \|^2 + 2 \| \mathbf{w} \|^2$.

2.1. Verify the Cauchy–Schwarz and triangle inequalities for the vectors $\mathbf{v} = (1, 2)^T$ and $\mathbf{w} = (1, -3)^T$ using (a) ♡ the dot product; (b) ♢ the weighted inner product $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + 2 v_2 w_2$; (c) the inner product (2.10).

Solution: (a) $| \mathbf{v} \cdot \mathbf{w} | = 5 \leq 7.0711 = \sqrt{5}\,\sqrt{10} = \| \mathbf{v} \| \, \| \mathbf{w} \|$.

2.2. Verify the Cauchy–Schwarz and triangle inequalities for each of the following pairs of vectors $\mathbf{v}, \mathbf{w}$, using the standard dot product, and then determine the angle between them: (a) ♡ $(1, 2)^T, (-1, 2)^T$, (b) ♢ $(1, -1, 0)^T, (-1, 0, 1)^T$, (c) $(1, -1, 1, 0)^T, (-2, 0, -1, 1)^T$.

Solution: (a) $| \mathbf{v}_1 \cdot \mathbf{v}_2 | = 3 \leq 5 = \sqrt{5}\,\sqrt{5} = \| \mathbf{v}_1 \| \, \| \mathbf{v}_2 \|$; angle: $\cos^{-1} \frac{3}{5} \approx .9273$.

2.4. ♡ Given an inner product on $\mathbb{R}^n$, define the corresponding (non-Euclidean) *angle* θ between two nonzero vectors $\mathbf{0} \neq \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ by the formula $\langle \mathbf{v}, \mathbf{w} \rangle = \| \mathbf{v} \| \, \| \mathbf{w} \| \, \cos \theta$. Prove that the *Law of Cosines* holds in general:

$$\| \mathbf{v} - \mathbf{w} \|^2 = \| \mathbf{v} \|^2 + \| \mathbf{w} \|^2 - 2 \| \mathbf{v} \| \, \| \mathbf{w} \| \, \cos \theta. \tag{2.31}$$

Solution:
$\| \mathbf{v} - \mathbf{w} \|^2 = \langle \mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w} \rangle = \| \mathbf{v} \|^2 - 2 \langle \mathbf{v}, \mathbf{w} \rangle + \| \mathbf{w} \|^2 = \| \mathbf{v} \|^2 + \| \mathbf{w} \|^2 - 2 \| \mathbf{v} \| \, \| \mathbf{w} \| \cos \theta$.

3.1. ♡ (a) Find $a \in \mathbb{R}$ such that $(2, a, -3)^T$ is orthogonal to $(-1, 3, -2)^T$.

(b) Is there any value of a for which $(2, a, -3)^T$ is parallel to $(-1, 3, -2)^T$?

Solution: (a) $a = -\frac{4}{3}$; (b) no.

3.2. ♡ Find all vectors in $\mathbb{R}^3$ that are orthogonal to both $(1,2,3)^T$ and $(-2,0,1)^T$.

Solution: All scalar multiples of $(2,-7,4)^T$.

3.5. Using the dot product, classify the following pairs of vectors in $\mathbb{R}^2$ as
(*i*) basis, (*ii*) orthogonal basis, and/or (*iii*) orthonormal basis:

$$(a)\heartsuit \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}; \ (b)\diamondsuit \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}; \ (c) \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}; \ (d)\heartsuit \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ -6 \end{pmatrix};$$

$$(e)\diamondsuit \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \end{pmatrix}; \ (f) \begin{pmatrix} \frac{3}{5} \\ \frac{4}{5} \end{pmatrix}, \begin{pmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{pmatrix}.$$

Solution: (*a*) Orthogonal basis; (*d*) basis.

3.6. Repeat Exercise 3.5, but use the weighted inner product $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + \frac{1}{9} v_2 w_2$ instead of the dot product.

Solution: (*a*) Basis; (*d*) orthogonal basis.

3.8. ♡ Suppose that $\mathbf{u}_1, \ldots, \mathbf{u}_n$ form an orthonormal basis of $\mathbb{R}^n$. Prove that the inner product between two vectors $\mathbf{v} = c_1 \mathbf{u}_1 + \cdots + c_n \mathbf{u}_n$ and $\mathbf{w} = d_1 \mathbf{u}_1 + \cdots + d_n \mathbf{u}_n$ is equal to the dot product of their coordinates: $\langle \mathbf{v}, \mathbf{w} \rangle = c_1 d_1 + \cdots + c_n d_n$.

Solution: Using bilinearity,

$$\langle \mathbf{v}, \mathbf{w} \rangle = \left\langle \sum_{i=1}^{n} c_i \mathbf{u}_i, \sum_{i=1}^{n} d_j \mathbf{u}_j \right\rangle = \sum_{i,j=1}^{n} c_i d_j \langle \mathbf{u}_i, \mathbf{u}_j \rangle = \sum_{i=1}^{n} c_i d_i. \qquad \blacksquare$$

4.1. Using the dot product on $\mathbb{R}^3$, given $\mathbf{v} = (1,1,1)^T$ find its orthogonal projection onto and distance to the following subspaces: (*a*) ♡ the line in the direction $\left(-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right)^T$; (*b*) the line spanned by $(2,-1,3)^T$; (*c*) ♢ the plane spanned by $(1,1,0)^T, (-2,2,1)^T$.

Solution: (*a*) projection: $\begin{pmatrix} -\frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}$, distance: $\dfrac{2\sqrt{2}}{3}$.

4.2. Redo Exercise 4.1 using the weighted inner product $\langle \mathbf{v}, \mathbf{w} \rangle = 2 v_1 w_1 + 2 v_2 w_2 + 3 v_3 w_3$.

Solution: (*a*) projection: $\begin{pmatrix} -\frac{3}{7} \\ \frac{3}{7} \\ \frac{3}{7} \end{pmatrix}$, distance: $2\sqrt{\dfrac{10}{7}}$.

5.1. Use the first version of the Gram–Schmidt process to determine an orthonormal basis for $\mathbb{R}^3$ with the dot product starting with the following sets of vectors:

$$(a)\heartsuit \begin{pmatrix} 1\\0\\1 \end{pmatrix}, \begin{pmatrix} 1\\1\\1 \end{pmatrix}, \begin{pmatrix} -1\\2\\1 \end{pmatrix}; \quad (b)\heartsuit \begin{pmatrix} 1\\1\\0 \end{pmatrix}, \begin{pmatrix} 0\\1\\-1 \end{pmatrix}, \begin{pmatrix} 1\\0\\-1 \end{pmatrix}; \quad (c)\diamondsuit \begin{pmatrix} 1\\2\\3 \end{pmatrix}, \begin{pmatrix} 4\\5\\0 \end{pmatrix}, \begin{pmatrix} 2\\3\\-1 \end{pmatrix}.$$

Solution: (a) $\frac{1}{\sqrt{2}}\,(1,0,1)^T,\ (0,1,0)^T,\ \frac{1}{\sqrt{2}}\,(-1,0,1)^T$;
(b) $\frac{1}{\sqrt{2}}\,(1,1,0)^T,\ \frac{1}{\sqrt{6}}\,(-1,1,-2)^T,\ \frac{1}{\sqrt{3}}\,(1,-1,-1)^T$.

5.2. Apply the Gram–Schmidt process to the following sets of vectors using the dot product on $\mathbb{R}^4$. Which produce an orthonormal basis?

$(a)\heartsuit$ $(1,0,1,0)^T, (0,1,0,-1)^T, (1,0,0,1)^T, (1,1,1,1)^T$;
(b) $(1,0,0,1)^T, (4,1,0,0)^T, (1,0,2,1)^T, (0,2,0,1)^T$;
$(c)\diamondsuit$ $(1,-1,0,1)^T, (0,-1,1,2)^T, (2,-1,-1,0)^T, (2,2,-2,1)^T$.

Solution: (a) $\left(\frac{1}{\sqrt{2}},0,\frac{1}{\sqrt{2}},0\right)^T,\ \left(0,\frac{1}{\sqrt{2}},0,-\frac{1}{\sqrt{2}}\right)^T,\ \left(\frac12,\frac12,-\frac12,\frac12\right)^T,\ \left(-\frac12,\frac12,\frac12,\frac12\right)^T$.

5.4. Use the Gram–Schmidt process to construct an orthonormal basis under the dot product for the following subspaces of $\mathbb{R}^3$: $(a)\heartsuit$ the plane spanned by $(0,2,1)^T, (1,-2,-1)^T$; $(b)\diamondsuit$ the plane defined by the equation $2x-y+3z=0$; (c) the set of all vectors orthogonal to $(1,-1,-2)^T$.

Solution: (a) $\left(0,\frac{2}{\sqrt{5}},\frac{1}{\sqrt{5}}\right)^T,\ (1,0,0)^T$.

5.5. Redo Exercises 5.1 and 5.4 using the weighted inner product

$$\langle \mathbf{v},\mathbf{w}\rangle = 3v_1w_1 + 2v_2w_2 + v_3w_3.$$

Solution: 5.1 (a) $\frac12\,(1,0,1)^T,\ \frac{1}{\sqrt{2}}\,(0,1,0)^T,\ \frac{1}{2\sqrt{3}}\,(-1,0,3)^T$;
(b) $\frac{1}{\sqrt{5}}\,(1,1,0)^T,\ \frac{1}{\sqrt{55}}\,(-2,3,-5)^T,\ \frac{1}{\sqrt{66}}\,(2,-3,-6)^T$.
5.4 (a) $\frac13\,(0,2,1)^T,\ \frac{1}{\sqrt{130}}\,(4,3,-8)^T$; (b) $\frac{1}{\sqrt{11}}\,(1,2,0)^T,\ \frac{1}{\sqrt{715}}\,(-12,9,11)^T$.

5.6.$\heartsuit$ Using the dot product on $\mathbb{R}^3$, find the orthogonal projection of the vector $(1,3,-1)^T$ onto the plane spanned by $(-1,2,1)^T, (2,1,-3)^T$ by first using the Gram–Schmidt process to construct an orthonormal basis.

Solution:
Orthonormal basis: $\left(-\frac{1}{\sqrt{6}},\frac{2}{\sqrt{6}},\frac{1}{\sqrt{6}}\right)^T,\ \left(\frac{3}{5\sqrt{2}},\frac{4}{5\sqrt{2}},-\frac{1}{\sqrt{2}}\right)^T$; projection: $\left(\frac{8}{15},\frac{44}{15},-\frac43\right)^T$.

6.1. Using the dot product on $\mathbb{R}^3$, find the orthogonal complement $V^\perp$ of the subspaces $V \subset \mathbb{R}^3$ spanned by the indicated vectors. What is the dimension of $V^\perp$ in each case?

$$(a)\heartsuit \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}, \quad (b)\heartsuit \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \quad (c) \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}, \quad (d)\diamondsuit \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

Solution:

(a) $V^\perp$ has basis $\begin{pmatrix} \frac{1}{3} \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{1}{3} \\ 0 \\ 1 \end{pmatrix}$, $\dim V^\perp = 2$; (b) $V^\perp$ has basis $\begin{pmatrix} -\frac{1}{2} \\ -\frac{5}{4} \\ 1 \end{pmatrix}$, $\dim V^\perp = 1$.

6.2. Use the dot product to decompose each of the following vectors with respect to the indicated subspace as $\mathbf{b} = \mathbf{p} + \mathbf{q}$, where $\mathbf{p} \in V$, $\mathbf{q} \in V^\perp$.

$$(a)\heartsuit\ \mathbf{b} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \ V = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \ \middle|\ 3x + 2y = 0 \right\}; \ (b)\diamondsuit\ \mathbf{b} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \ V = \operatorname{span} \left\{ \begin{pmatrix} -3 \\ 1 \end{pmatrix} \right\};$$

$$(c)\ \mathbf{b} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \ V = \{ x - y + z = 0 \}; \ (d)\ \mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \ V = \operatorname{span} \left\{ \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Solution: (a) $\mathbf{p} = \begin{pmatrix} -\frac{6}{13} \\ \frac{9}{13} \end{pmatrix}$, $\mathbf{q} = \begin{pmatrix} \frac{6}{13} \\ \frac{4}{13} \end{pmatrix}$.

6.3. Find an orthonormal basis under the dot product for the orthogonal complement of the following subspaces of $\mathbb{R}^3$: $(a)\heartsuit$ the plane $3x + 4y - 5z = 0$; (b) the plane spanned by $(1, -1, 3)^T, (2, 0, -1)^T$; $(c)\diamondsuit$ the line in the direction $(-2, 1, 3)^T$.

Solution: (a) $\dfrac{1}{5\sqrt{2}} \begin{pmatrix} 3 \\ 4 \\ -5 \end{pmatrix}$.

6.5. $\heartsuit$ Prove that if $V_1 \subset V_2 \subset \mathbb{R}^n$ are subspaces, then $V_1^\perp \supset V_2^\perp$.

Solution: If $\mathbf{z} \in V_2^\perp$ then $\langle \mathbf{z}, \mathbf{w} \rangle = 0$ for every $\mathbf{w} \in V_2$. In particular, since $V_1 \subset V_2$, every $\mathbf{w} \in V_1$ belongs to V_2, and hence $\mathbf{z}$ is orthogonal to every vector $\mathbf{w} \in V_1$. Thus, $\mathbf{z} \in V_1^\perp$, proving $V_2^\perp \subset V_1^\perp$. ■

7.1. Compute the 1, 2, 3, and ∞ norms of the following vectors, and then verify the triangle inequality in each case.

$$(a)\heartsuit \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}; \ (b) \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \end{pmatrix}; \ (c) \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}; \ (d)\diamondsuit \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ -3 \end{pmatrix}.$$

Solution: (a)
$$\begin{aligned}
\| \mathbf{v} + \mathbf{w} \|_1 &= 2 \le 1 + 1 = \| \mathbf{v} \|_1 + \| \mathbf{w} \|_1, \\
\| \mathbf{v} + \mathbf{w} \|_2 &= \sqrt{2} \le 1 + 1 = \| \mathbf{v} \|_2 + \| \mathbf{w} \|_2, \\
\| \mathbf{v} + \mathbf{w} \|_3 &= \sqrt[3]{2} \le 1 + 1 = \| \mathbf{v} \|_3 + \| \mathbf{w} \|_3, \\
\| \mathbf{v} + \mathbf{w} \|_\infty &= 1 \le 1 + 1 = \| \mathbf{v} \|_\infty + \| \mathbf{w} \|_\infty.
\end{aligned}$$

7.2. Find a unit vector in the same direction as $\mathbf{v} = (1, 2, -3)^T$ for (*a*) ♡ the Euclidean norm, (*b*) ♢ the weighted norm $\|\mathbf{v}\|^2 = 2v_1^2 + v_2^2 + \frac{1}{3}v_3^2$, (*c*) ♡ the 1 norm, (*d*) the ∞ norm.

Solution: (*a*) $\left(\frac{1}{\sqrt{14}}, \frac{2}{\sqrt{14}}, -\frac{3}{\sqrt{14}} \right)^T$; (*c*) $\left(\frac{1}{6}, \frac{1}{3}, -\frac{1}{2} \right)^T$.

7.3. Which two of the vectors $\mathbf{u} = (-2, 2, 1)^T$, $\mathbf{v} = (1, 4, 1)^T$, $\mathbf{w} = (0, 0, -1)^T$ are closest in distance for (*a*) ♡ the Euclidean norm? (*b*) ♢ the 1 norm? (*c*) the ∞ norm?

Solution: (*a*) $\|\mathbf{u} - \mathbf{v}\|_2 = \sqrt{13}$, $\|\mathbf{u} - \mathbf{w}\|_2 = \sqrt{12}$, $\|\mathbf{v} - \mathbf{w}\|_2 = \sqrt{21}$, so $\mathbf{u}, \mathbf{w}$ are closest.

7.5. Prove that the following formulas define norms on $\mathbb{R}^2$:
(*a*) ♡ $\|\mathbf{v}\| = \sqrt{2v_1^2 + 3v_2^2}$, (*b*) $\|\mathbf{v}\| = \sqrt{2v_1^2 - v_1 v_2 + 2v_2^2}$, (*c*) ♡ $\|\mathbf{v}\| = 2|v_1| + |v_2|$, (*d*) $\|\mathbf{v}\| = \max\{2|v_1|, |v_2|\}$, (*e*) ♢ $\|\mathbf{v}\| = \max\{|v_1 - v_2|, |v_1 + v_2|\}$.

Solution: (*a*) Comes from weighted inner product $\langle \mathbf{v}, \mathbf{w} \rangle = 2v_1 w_1 + 3v_2 w_2$.
(*c*) Clearly positive; $\|c\mathbf{v}\| = 2|cv_1| + |cv_2| = |c|\left(2|v_1| + |v_2|\right) = |c|\,\|\mathbf{v}\|$;
$\|\mathbf{v} + \mathbf{w}\| = 2|v_1 + w_1| + |v_2 + w_2| \leq 2|v_1| + |v_2| + 2|w_1| + |w_2| = \|\mathbf{v}\| + \|\mathbf{w}\|$.

7.6. Which of the following formulas define norms on $\mathbb{R}^3$? (*a*) ♡ $\|\mathbf{v}\| = \sqrt{2v_1^2 + v_2^2 + 3v_3^2}$, (*b*) ♡ $\|\mathbf{v}\| = \sqrt{v_1^2 + 2v_1 v_2 + v_2^2 + v_3^2}$, (*c*) ♢ $\|\mathbf{v}\| = \max\{|v_1|, |v_2|, |v_3|\}$, (*d*) $\|\mathbf{v}\| = |v_1 - v_2| + |v_2 - v_3| + |v_3 - v_1|$, (*e*) $\|\mathbf{v}\| = |v_1| + \max\{|v_2|, |v_3|\}$.

Solution: (*a*) Norm; (*b*) not a norm since, for instance, $\|(1, -1, 0)^T\| = 0$.

7.7. ♡ Prove that any norm on $\mathbb{R}^n$ satisfies the *reverse triangle inequality*

$$\|\mathbf{x} + \mathbf{y}\| \geq \big|\,\|\mathbf{x}\| - \|\mathbf{y}\|\,\big| \qquad \text{for all} \qquad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n. \tag{2.79}$$

Solution: The regular triangle inequality yields $\|\mathbf{x}\| = \|\mathbf{x} + \mathbf{y} - \mathbf{y}\| \leq \|\mathbf{x} + \mathbf{y}\| + \|\mathbf{y}\|$, which implies $\|\mathbf{x} + \mathbf{y}\| \geq \|\mathbf{x}\| - \|\mathbf{y}\|$. Swapping $\mathbf{x}$ and $\mathbf{y}$, we also have $\|\mathbf{x} + \mathbf{y}\| \geq \|\mathbf{y}\| - \|\mathbf{x}\|$. Combining these proves (2.79). ■

7.10. ♡ *True or false*: If $\|\mathbf{v} + \mathbf{w}\| = \|\mathbf{v}\| + \|\mathbf{w}\|$, then $\mathbf{v}, \mathbf{w}$ are parallel vectors.

Solution: True for an inner product norm, but false in general. For example,
$$\|\mathbf{e}_1 + \mathbf{e}_2\|_1 = 2 = \|\mathbf{e}_1\|_1 + \|\mathbf{e}_2\|_1.$$

7.11. ♡ How many unit vectors are parallel to a given vector $\mathbf{v} \neq \mathbf{0}$? (*a*) 0, (*b*) 1, (*c*) 2, (*d*) 3, (*e*) ∞, (*f*) depends on the norm. Explain your answer.

Solution: 2 vectors, namely $\mathbf{u} = \mathbf{v}/\|\mathbf{v}\|$ and $-\mathbf{u} = -\mathbf{v}/\|\mathbf{v}\|$.

7.14. Check the validity of the inequalities (2.74) for the particular vectors
(*a*) ♡ $(1, -1)^T$, (*b*) ♢ $(1, 2, 3)^T$, (*c*) $(1, 1, 1, 1)^T$.

Solution: (*a*) $\|\mathbf{v}\|_2 = \sqrt{2}$, $\|\mathbf{v}\|_\infty = 1$, and $\frac{1}{\sqrt{2}}\sqrt{2} \leq 1 \leq \sqrt{2}$.

7.17. Compute the cosine distance between the pairs of vectors in Exercise 7.1.

Solution: (*a*) 1.

Chapter 3

Matrices

1.2. Let $A = \begin{pmatrix} 1 & -1 & 3 \\ -1 & 4 & -2 \\ 3 & 0 & 6 \end{pmatrix}$, $B = \begin{pmatrix} -6 & 0 & 3 \\ 4 & 2 & -1 \end{pmatrix}$, $C = \begin{pmatrix} 2 & 3 \\ -3 & -4 \\ 1 & 2 \end{pmatrix}$.

Compute the indicated combinations where possible. $(a)\heartsuit$ $3A - B$, (b) AB, $(c)\heartsuit$ BA, $(d)\heartsuit$ $(A+B)C$, (e) $A + BC$, $(f)\diamondsuit$ $A + 2CB$, (g) $A^2 - 3A + \mathrm{I}$, (h) $(B - \mathrm{I})(C + \mathrm{I})$.

Solution: (a) Undefined, (c) $\begin{pmatrix} 3 & 6 & 0 \\ -1 & 4 & 2 \end{pmatrix}$, (d) undefined.

1.3. Which of the following pairs of matrices commute under matrix multiplication?

$(a)\heartsuit$ $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$, $\begin{pmatrix} 4 & 3 \end{pmatrix}$, (b) $\begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$, $\begin{pmatrix} 2 & 3 \\ 5 & 0 \end{pmatrix}$, $(c)\heartsuit$ $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, $\begin{pmatrix} 3 & -2 \\ -2 & 3 \end{pmatrix}$,

$(d)\diamondsuit$ $\begin{pmatrix} 3 & -1 \\ 0 & 2 \\ 1 & 4 \end{pmatrix}$, $\begin{pmatrix} 4 & 2 & -2 \\ 5 & 2 & 4 \end{pmatrix}$, (e) $\begin{pmatrix} 3 & 0 & -1 \\ -2 & -1 & 2 \\ 2 & 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 2 & 0 & -1 \\ 1 & 1 & -1 \\ 2 & 0 & -1 \end{pmatrix}$.

Solution: (a) Do not commute; (c) commute.

1.6. $\heartsuit$ Find a nonzero matrix $A \neq \mathrm{O}$ such that $A^2 = \mathrm{O}$.

Solution: For example, $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$.

1.8. $(a)\heartsuit$ Let A be an $m \times n$ matrix. Let $\mathbf{e}_j \in \mathbb{R}^n$ denote the j-th standard basis vector. Explain why the product $A\mathbf{e}_j$ equals the j-th column of A. $(b)\diamondsuit$ Similarly, let $\widehat{\mathbf{e}}_i \in \mathbb{R}^m$ be the i-th standard basis vector. Explain why the triple product $\widehat{\mathbf{e}}_i^T A\mathbf{e}_j = a_{ij}$ equals the (i,j) entry of the matrix A.

Solution: (a) The i-th entry of $A\mathbf{e}_j$ is the product of the i-th row of A with $\mathbf{e}_j$. Since all the entries in $\mathbf{e}_j$ are zero except the j-th entry the product will be equal to a_{ij}, i.e., the (i,j) entry of A, which are the entries of its j-th column. ■

1.9. ♡ Prove that $A\mathbf{v} = \mathbf{0}$ for every vector $\mathbf{v}$ (with the appropriate number of entries) if and only if $A = \mathrm{O}$ is the zero matrix.

Solution: Let $\mathbf{v} = \mathbf{e}_j$. Then $A\mathbf{v}$ is the same as the j-th column of A. Thus, the hypothesis implies all columns of A are $\mathbf{0}$ and hence $A = \mathrm{O}$. ■

1.13. Write out the following diagonal matrices: (*a*) ♡ $\operatorname{diag}(1, 0, -1)$, (*b*) $\operatorname{diag}(2, -2, 3, -3)$.

Solution: (a) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$.

1.15. The *trace* of a square matrix $A \in \mathcal{M}_{n\times n}$ is defined to be the sum of its diagonal entries:

$$\operatorname{tr} A = a_{11} + a_{22} + \cdots + a_{nn}. \tag{3.12}$$

Let A, B, C be $n \times n$ matrices. Prove that the trace satisfies the following identities: (*a*) ♡ $\operatorname{tr}(A + B) = \operatorname{tr} A + \operatorname{tr} B$; (*b*) ♡ $\operatorname{tr}(AB) = \operatorname{tr}(BA)$; (*c*) ♢ $\operatorname{tr}(ABC) = \operatorname{tr}(CAB) = \operatorname{tr}(BCA)$. On the other hand, find an example where $\operatorname{tr}(ABC) \neq \operatorname{tr}(ACB)$. (*d*) Is part (*b*) valid if A has size $m \times n$ and B has size $n \times m$?

Solution: (a) $\operatorname{tr}(A + B) = \displaystyle\sum_{i=1}^{n} (a_{ii} + b_{ii}) = \sum_{i=1}^{n} a_{ii} + \sum_{i=1}^{n} b_{ii} = \operatorname{tr} A + \operatorname{tr} B$. ■

(*b*) The diagonal entries of AB are $\displaystyle\sum_{j=1}^{n} a_{ij} b_{ji}$, so $\operatorname{tr}(AB) = \displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} b_{ji}$; the diagonal entries of BA are $\displaystyle\sum_{i=1}^{n} b_{ji} a_{ij}$, so $\operatorname{tr}(BA) = \displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ji} a_{ij}$. These are clearly equal. ■

1.17. The naïve way to "multiply" matrices is known as the *Hadamard product*, and is occasionally useful. More specifically, given two $m \times n$ matrices A, B, necessarily of the *same* size, their Hadamard product is the $m \times n$ matrix $C = A \circ B$ whose (i, j) entry is merely the product of the (i, j) entries of A and B, so $c_{ij} = a_{ij} b_{ij}$.

(*a*) ♡ Prove that the Hadamard product is commutative: $A \circ B = B \circ A$.
(*b*) Which of the matrix arithmetic properties does the Hadamard product satisfy?
(*c*) ♡ What is the multiplicative identity for the Hadamard product?
(*d*) Let $D = \operatorname{diag} \mathbf{d}$ be a diagonal matrix. Show that $D\mathbf{x} = \mathbf{d} \circ \mathbf{x}$.
(*e*) ♢ Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$. Prove the Hadamard product vector identities

$$(i)\ (\mathbf{x} \circ \mathbf{y}) \cdot \mathbf{z} = (\mathbf{x} \circ \mathbf{z}) \cdot \mathbf{y}, \qquad (ii)\ (\mathbf{x}\mathbf{x}^T) \circ (\mathbf{y}\mathbf{y}^T) = (\mathbf{x} \circ \mathbf{y})(\mathbf{x} \circ \mathbf{y})^T.$$

Solution:

(a) The (i, j) entry of $A \circ B$ is $a_{ij} b_{ij} = b_{ij} a_{ij}$, which equals the (i, j) entry of $B \circ A$. ■

(c) The multiplicative identity for the Hadamard product is the $m \times n$ matrix E that has all 1's as entries: $A \circ E = A = E \circ A$.

2.1. Write down the transpose of the following matrices:

$$(a)\heartsuit \begin{pmatrix} 1 \\ 5 \end{pmatrix}, \quad (b) \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \quad (c)\heartsuit \begin{pmatrix} 1 & 2 & -1 \\ 2 & 0 & 2 \end{pmatrix}, \quad (d) \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}, \quad (e)\diamondsuit \begin{pmatrix} 1 & 2 & -1 \\ 0 & 3 & 2 \\ 1 & 1 & 5 \end{pmatrix}.$$

Solution: (a) $\begin{pmatrix} 1 & 5 \end{pmatrix}$, (c) $\begin{pmatrix} 1 & 2 \\ 2 & 0 \\ -1 & 2 \end{pmatrix}$.

2.4. $\heartsuit$ Let A be a square matrix. Prove that $A + A^T$ is symmetric.

Solution: $(A + A^T)^T = A^T + (A^T)^T = A^T + A = A + A^T$. ■

2.6. If $\mathbf{v}, \mathbf{w}$ are column vectors with the same number of entries, does

$$(a)\heartsuit \ \mathbf{v}^T\mathbf{w} = \mathbf{w}^T\mathbf{v}? \qquad (b)\diamondsuit \ \mathbf{v}\,\mathbf{w}^T = \mathbf{w}\,\mathbf{v}^T?$$

Solution: (a) Yes, since the expression is a scalar, so $\mathbf{v}^T\mathbf{w} = (\mathbf{v}^T\mathbf{w})^T = \mathbf{w}^T(\mathbf{v}^T)^T = \mathbf{w}^T\mathbf{v}$.

2.8. $\heartsuit$ Let $A = (\,\mathbf{v}_1 \ \ldots \ \mathbf{v}_n\,)$ be an $m \times n$ matrix with the indicated columns. Prove that the trace (see Exercise 1.15) of the symmetric matrix A^TA equals the sum of the squared Euclidean norms of the columns of A, i.e., $\operatorname{tr}(A^TA) = \|\,\mathbf{v}_1\,\|_2^2 + \cdots + \|\,\mathbf{v}_n\,\|_2^2$.

Solution: The diagonal entries of A^TA are $\mathbf{v}_i^T\mathbf{v}_i = \|\,\mathbf{v}_i\,\|_2^2$. Summing them from $i = 1, \ldots, n$ produces the formula. ■

2.9. Suppose R, S are symmetric matrices. Prove that (a) $\heartsuit$ their sum $R + S$ is symmetric; (b) $\diamondsuit$ their product RS is symmetric if and only if R and S commute: $RS = SR$.

Solution: (a) $(R + S)^T = R^T + S^T = R + S$.

3.1. For each of the following linear systems, write down the coefficient matrix A and the vectors $\mathbf{x}$ and $\mathbf{b}$.

$$(a)\heartsuit \ \begin{aligned} x - y &= 7, \\ x + 2y &= 3; \end{aligned} \quad (b) \ \begin{aligned} 6u + v &= 5, \\ 3u - 2v &= 5; \end{aligned} \quad (c)\diamondsuit \ \begin{aligned} p + 3q - 3r &= 0, \\ q - r &= 1, \\ 2p - q + 3r &= 3, \\ 2p - 5r &= -1; \end{aligned} \quad (d)\heartsuit \ \begin{aligned} 2u - v + 2w &= 2, \\ -u - v &= 1, \\ 3u - 2w &= 1. \end{aligned}$$

Solution: (a) $A = \begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 7 \\ 3 \end{pmatrix}$

(d) $A = \begin{pmatrix} 2 & -1 & 2 \\ -1 & -1 & 0 \\ 3 & 0 & -2 \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} u \\ v \\ w \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$.

3.2. Write out and solve the linear systems corresponding to the indicated matrix, vector of unknowns, and right-hand side. (a) $\heartsuit$ $A = \begin{pmatrix} 1 & -1 \\ 2 & 3 \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} -1 \\ -3 \end{pmatrix}$;

(b) $\diamondsuit$ $A = \begin{pmatrix} 0 & -4 \\ 5 & 1 \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} c \\ d \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$; (c) $\heartsuit$ $A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} u \\ v \\ w \end{pmatrix}$,

$\mathbf{b} = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$; (d) $A = \begin{pmatrix} 3 & 0 & -1 \\ -2 & -1 & 0 \\ 0 & -3 & 0 \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.

Solution: (a) $x - y = -1,\ \ 2x + 3y = -3$. The solution is $x = -\frac{6}{5},\ y = -\frac{1}{5}$.
(c) $u + w = -1,\ \ u + v = -1,\ \ v + w = 2$. The solution is $u = -2,\ v = 1,\ w = 1$.

3.3. Write out the linear system that determines whether the following sets of vectors are linearly independent or dependent. Then determine which of the two possibilities holds.

$$(a)\heartsuit \begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}2\\1\end{pmatrix}, \quad (b) \begin{pmatrix}1\\3\end{pmatrix}, \begin{pmatrix}-2\\-6\end{pmatrix}, \quad (c)\heartsuit \begin{pmatrix}2\\1\end{pmatrix}, \begin{pmatrix}-1\\3\end{pmatrix}, \begin{pmatrix}5\\2\end{pmatrix}, \quad (d) \begin{pmatrix}1\\3\\-2\end{pmatrix}, \begin{pmatrix}0\\2\\-1\end{pmatrix},$$

$$(e)\diamondsuit \begin{pmatrix}0\\1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\\0\end{pmatrix}, \begin{pmatrix}3\\-1\\2\end{pmatrix}, \quad (f) \begin{pmatrix}2\\1\\3\end{pmatrix}, \begin{pmatrix}1\\-2\\1\end{pmatrix}, \begin{pmatrix}2\\-3\\0\end{pmatrix}, \begin{pmatrix}0\\-1\\4\end{pmatrix}.$$

Solution: (a) $A = \begin{pmatrix}1 & -1\\1 & 2\end{pmatrix},\ \mathbf{x} = \begin{pmatrix}x_1\\x_2\end{pmatrix},\ \mathbf{b} = \begin{pmatrix}0\\0\end{pmatrix}$, linearly independent;

(c) $A = \begin{pmatrix}2 & -1 & 5\\1 & 3 & 2\end{pmatrix},\ \mathbf{x} = \begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix},\ \mathbf{b} = \begin{pmatrix}0\\0\end{pmatrix}$, linearly dependent.

3.4. For each of the corresponding sets of vectors in Exercise 3.3, write out the linear system that determines whether the indicated vector lies in their span. Then determine whether or not this holds.

$$(a)\heartsuit \begin{pmatrix}1\\0\end{pmatrix}, \quad (b) \begin{pmatrix}2\\1\end{pmatrix}, \quad (c) \begin{pmatrix}-1\\-1\end{pmatrix}, \quad (d)\heartsuit \begin{pmatrix}1\\0\\0\end{pmatrix}, \quad (e)\diamondsuit \begin{pmatrix}1\\1\\1\end{pmatrix}, \quad (f) \begin{pmatrix}2\\-2\\4\end{pmatrix}.$$

Solution: (a) $A = \begin{pmatrix}1 & -1\\1 & 2\end{pmatrix},\ \mathbf{x} = \begin{pmatrix}x_1\\x_2\end{pmatrix},\ \mathbf{b} = \begin{pmatrix}1\\0\end{pmatrix}$, in span;

(d) $A = \begin{pmatrix}1 & 0\\3 & 2\\-2 & -1\end{pmatrix},\ \mathbf{x} = \begin{pmatrix}x_1\\x_2\end{pmatrix},\ \mathbf{b} = \begin{pmatrix}1\\0\\0\end{pmatrix}$, not in span.

4.1. Find a basis, if it exists, of the image and the kernel of the following matrices:

$$(a)\heartsuit \begin{pmatrix}2 & -1 & 5\end{pmatrix}, \quad (b) \begin{pmatrix}8 & -4\\-6 & 3\end{pmatrix}, \quad (c)\heartsuit \begin{pmatrix}1 & -1 & 2\\-2 & 2 & -4\end{pmatrix}, \quad (d)\diamondsuit \begin{pmatrix}1 & 2 & 3\\0 & 4 & 5\\0 & 0 & 6\end{pmatrix}.$$

Solution: (a) Image: (1); kernel: $\begin{pmatrix}1\\2\\0\end{pmatrix}, \begin{pmatrix}0\\5\\1\end{pmatrix}$. (c) Image: $\begin{pmatrix}1\\-1\end{pmatrix}$; kernel $\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}\frac{1}{2}\\0\\1\end{pmatrix}$.

4.5. $\heartsuit$ *True or false*: If A is a square matrix, then $\ker A \cap \operatorname{img} A = \{\mathbf{0}\}$.

Solution: False. For example, if $A = \begin{pmatrix}1 & 1\\-1 & -1\end{pmatrix}$ then $\begin{pmatrix}1\\-1\end{pmatrix}$ is in both $\ker A$ and $\operatorname{img} A$.

4.8. $\heartsuit$ *True or false*: If $\ker A = \ker B$, then $\operatorname{rank} A = \operatorname{rank} B$.

Solution: True. If $\ker A = \ker B \subset \mathbb{R}^n$, then both matrices have n columns, and so
$$n - \operatorname{rank} A = \dim\ker A = \dim\ker B = n - \operatorname{rank} B.$$

5.2. ♡ Let $A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 5 & -1 \\ 1 & 3 & 2 \end{pmatrix}$. Given that $\mathbf{x}_1^\star = \begin{pmatrix} 5 \\ -1 \\ 2 \end{pmatrix}$ solves $A\,\mathbf{x} = \mathbf{b}_1 = \begin{pmatrix} 1 \\ 3 \\ 6 \end{pmatrix}$ and $\mathbf{x}_2^\star = \begin{pmatrix} -11 \\ 5 \\ -1 \end{pmatrix}$ solves $A\,\mathbf{x} = \mathbf{b}_2 = \begin{pmatrix} 0 \\ 4 \\ 2 \end{pmatrix}$, find a solution to $A\,\mathbf{x} = 2\,\mathbf{b}_1 + \mathbf{b}_2 = \begin{pmatrix} 2 \\ 10 \\ 14 \end{pmatrix}$.

Solution: $\mathbf{x}^\star = 2\,\mathbf{x}_1^\star + \mathbf{x}_2^\star = (\,-1, 3, 3\,)^T$.

6.1. Verify by direct multiplication that the following matrices are inverses:

$$(a)\,♡\ \begin{pmatrix} 2 & 3 \\ -1 & -1 \end{pmatrix},\ \begin{pmatrix} -1 & -3 \\ 1 & 2 \end{pmatrix};\quad (b)\ \begin{pmatrix} 2 & 1 & 1 \\ 3 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix},\ \begin{pmatrix} 3 & -1 & -1 \\ -4 & 2 & 1 \\ -1 & 0 & 1 \end{pmatrix}.$$

Solution: (a) $\begin{pmatrix} 2 & 3 \\ -1 & -1 \end{pmatrix}\begin{pmatrix} -1 & -3 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & -3 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ -1 & -1 \end{pmatrix}$.

6.5. ♡ Prove that a diagonal matrix $D = \operatorname{diag}(d_1, \ldots, d_n)$ is invertible if and only if all its diagonal entries are nonzero, in which case $D^{-1} = \operatorname{diag}(1/d_1, \ldots, 1/d_n)$.

Solution: If all the diagonal entries are nonzero, then $D^{-1}D = \mathrm{I}$. On the other hand, if one of diagonal entries is zero, then all the entries in that row are zero, and so D is not invertible. ■

6.7. ♡ (a) Prove that the inverse transpose operation (3.53) respects matrix multiplication: $(A\,B)^{-T} = A^{-T}B^{-T}$. (b) Verify this identity for $A = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$.

Solution: (a) $(A\,B)^{-T} = ((A\,B)^T)^{-1} = (B^T A^T)^{-1} = (A^T)^{-1}(B^T)^{-1} = A^{-T}B^{-T}$. ■

(b) $A\,B = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$, so $(A\,B)^{-T} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$,

while $A^{-T} = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$, $B^{-T} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$, so $A^{-T}B^{-T} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$. ■

7.1. ♡ (a) Show that the function $R\colon \mathbb{R}^2 \to \mathbb{R}^2$ that rotates vectors in the plane by 90° is linear and find its matrix representative.

(b) Answer the same question for rotation by a specified angle θ.

Solution: Rotations preserve parallelograms (vector addition) and stretching of vectors (scalar multiplication) and hence define linear maps. (a) $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$; (b) $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

7.5. ♡ Prove that an affine function maps parallel lines to parallel lines.

Solution: Given parallel lines parametrized by $\mathbf{c} + t\,\mathbf{v}$ and $\mathbf{d} + s\,\mathbf{v}$, and an affine map (3.63), the image lines are parametrized by $(A\,\mathbf{c} + \mathbf{b}) + t\,A\,\mathbf{v}$ and $(A\,\mathbf{d} + \mathbf{b}) + s\,A\,\mathbf{v}$, which are also parallel. ■

7.7. Suppose we identify $\mathcal{M}_{m\times n} \simeq \mathbb{R}^{mn}$. (a) ♡ Show that, for a fixed $k \times m$ matrix B, matrix multiplication $L[A] = BA$ defines a linear function $L\colon \mathbb{R}^{mn} \to \mathbb{R}^{km}$. (b) Show that, similarly, the trace $\operatorname{tr} A$ of a square matrix $A \in \mathcal{M}_{n\times n}$ defines a linear function $\operatorname{tr}\colon \mathbb{R}^{n^2} \to \mathbb{R}$.

Solution: (a) For any scalars c, d and matrices A, C, we have

$$L[cA + dC] = B(cA + dC) = cBA + dBC = cL[A] + dL[C],$$

proving linearity. ■

Chapter 4

How Matrices Interact with Inner Products and Norms

1.1. Are the following matrices are positive definite? In the positive definite cases, write down its Cholesky factorization and the formula for the associated inner product.

$$(a)\heartsuit \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \quad (b)\heartsuit \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad (c)\diamondsuit \begin{pmatrix} 1 & -1 \\ -1 & 3 \end{pmatrix}, \quad (d) \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 1 \end{pmatrix},$$

$$(e)\heartsuit \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & -2 \\ 1 & -2 & 4 \end{pmatrix}, \quad (f) \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}, \quad (g)\diamondsuit \begin{pmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}.$$

Solution: (a) Positive definite; (b) not positive definite; (e) not positive definite.

1.3. $\heartsuit$ Let $C = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$. Prove that the associated quadratic form $q(\mathbf{x}) = \mathbf{x}^T C\,\mathbf{x}$ is indefinite by finding a point $\mathbf{x}^+$ where $q(\mathbf{x}^+) > 0$ and a point $\mathbf{x}^-$ where $q(\mathbf{x}^-) < 0$.

Solution: For instance, $q(1,0) = 1$, while $q(2,-1) = -1$. ■

1.5. $(a)\heartsuit$ Prove that the sum of two positive definite matrices is positive definite.

(b) More generally, prove that the sum of a positive definite matrix and a positive semidefinite matrix is positive definite.

$(c)\diamondsuit$ Can the sum of two positive semidefinite matrices be positive definite?

(d) Give an example of two matrices that are not positive definite or semidefinite, but whose sum is positive definite.

Solution: (a) If $H, K > 0$, then $\mathbf{x}^T H\mathbf{x} > 0$ and $\mathbf{x}^T K\,\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$, hence

$$\mathbf{x}^T (H+K)\mathbf{x} = \mathbf{x}^T H\,\mathbf{x} + \mathbf{x}^T K\,\mathbf{x} > 0 \quad \text{for all } \mathbf{x} \neq \mathbf{0}.$$ ■

1.9. ♡ (a) Show that every diagonal entry of a positive definite matrix must be strictly positive. (b) Write down a symmetric matrix with all positive diagonal entries that is not positive definite. (c) Find a nonzero matrix with one or more zero diagonal entries that is positive semidefinite.

Solution: (a) $c_{ii} = \mathbf{e}_i^T C \mathbf{e}_i > 0$. ■ (b) For example, $C = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ is not positive definite or even semidefinite. (c) For example, $C = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

1.13. ♡ Let A be an $n \times n$ matrix. Prove that $\mathbf{x}^T A \mathbf{x} = \mathbf{x}^T S \mathbf{x}$, where $S = \frac{1}{2}(A + A^T)$ is a symmetric matrix. Therefore, we do not lose any generality by restricting our discussion to quadratic forms that are constructed from symmetric matrices.

Solution: Since $q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$ is a scalar $q(\mathbf{x}) = q(\mathbf{x})^T = (\mathbf{x}^T A \mathbf{x})^T = \mathbf{x}^T A^T \mathbf{x}$, and hence

$$q(\mathbf{x}) = \tfrac{1}{2}(\mathbf{x}^T A \mathbf{x} + \mathbf{x}^T A^T \mathbf{x}) = \mathbf{x}^T S \mathbf{x}. \qquad ■$$

2.1. Find the Gram matrix corresponding to each of the following sets of vectors using the Euclidean dot product on $\mathbb{R}^n$. Which are positive definite? (a) ♡ $\begin{pmatrix} -1 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}$,

(b) ♢ $\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix}$, (c) ♡ $\begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \\ 2 \end{pmatrix}$, (d) ♢ $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$,

(e) ♡ $\begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$, (f) $\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \\ 1 \end{pmatrix}$, (g) $\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ -4 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \\ -1 \\ -2 \end{pmatrix}$.

Solution: (a) $\begin{pmatrix} 10 & 6 \\ 6 & 4 \end{pmatrix}$; positive definite. (c) $\begin{pmatrix} 6 & -8 \\ -8 & 13 \end{pmatrix}$; positive definite. (e) $\begin{pmatrix} 9 & 6 & 3 \\ 6 & 6 & 0 \\ 3 & 0 & 3 \end{pmatrix}$; positive semidefinite.

2.2. Recompute the Gram matrices for cases $(c\text{–}e)$ in the previous exercise using the weighted inner product $\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + 2 x_2 y_2 + 3 x_3 y_3$. Does this change their positive definiteness?

Solution: (c) $\begin{pmatrix} 9 & -12 \\ -12 & 21 \end{pmatrix}$; (e) $\begin{pmatrix} 21 & 12 & 9 \\ 12 & 9 & 3 \\ 9 & 3 & 6 \end{pmatrix}$.

Positive definiteness doesn't change, since it only depends upon the linear independence of the vectors.

2.3. Express the following as Gram matrices or explain why this is not possible.

$(a)\heartsuit \begin{pmatrix} 2 & 3 \\ 3 & 4 \end{pmatrix}$, $(b)\heartsuit \begin{pmatrix} 4 & -1 \\ -1 & 1 \end{pmatrix}$, $(c)\diamondsuit \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix}$, $(d) \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$, $(e)\diamondsuit \begin{pmatrix} 9 & 3 & 3 \\ 3 & 2 & 2 \\ 3 & 2 & 6 \end{pmatrix}$.

Solution: (a) Not positive (semi)definite; $(b) \begin{pmatrix} 2 & 0 \\ -1/2 & \sqrt{3}/2 \end{pmatrix}\begin{pmatrix} 2 & -1/2 \\ 0 & \sqrt{3}/2 \end{pmatrix}$.

2.5. $\heartsuit$ (a) Prove that if C is a positive definite matrix, then C^2 is also positive definite.
(b) More generally, if S is symmetric and nonsingular, then S^2 is positive definite.

Solution: (a) is a special case of (b) since positive definite matrices are symmetric.
(b) By Theorem 4.12 if S is *any* symmetric matrix, then $S^T S = S^2$ is always positive semidefinite, and positive definite if and only if $\ker S = \{\mathbf{0}\}$, i.e., S is nonsingular. In particular, if $S = H > 0$, then $\ker H = \{\mathbf{0}\}$ and so $H^2 > 0$. ■

3.1. Choose one from the following list of inner products on $\mathbb{R}^3$ for both the domain and codomain, and find the adjoint of $A = \begin{pmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 2 \end{pmatrix}$: $(a)\heartsuit$ the Euclidean dot product; $(b)\heartsuit$ the weighted inner product $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + 2 v_2 w_2 + 3 v_3 w_3$; $(c)\diamondsuit$ the inner product $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T C \mathbf{w}$ defined by the positive definite matrix $C = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}$.

Solution: $(a) \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 2 \end{pmatrix}$, $(b) \begin{pmatrix} 1 & -2 & 0 \\ \frac{1}{2} & 0 & -\frac{3}{2} \\ 0 & \frac{2}{3} & 2 \end{pmatrix}$,.

3.4. $\heartsuit$ Consider the weighted inner product $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + \frac{1}{2} v_2 w_2 + \frac{1}{3} v_3 w_3$ on $\mathbb{R}^3$.
(a) What are the conditions on the entries of a 3×3 matrix A in order that it be self-adjoint?
(b) Write down an example of a non-diagonal self-adjoint matrix.

Solution: (a) $a_{12} = \frac{1}{2} a_{21}$, $a_{13} = \frac{1}{3} a_{31}$, $\frac{1}{2} a_{23} = \frac{1}{3} a_{32}$, $(b) \begin{pmatrix} 0 & 1 & 1 \\ 2 & 1 & 2 \\ 3 & 3 & 2 \end{pmatrix}$.

3.6. Prove the following adjoint identities: $(a)\heartsuit$ $(A+B)^* = A^* + B^*$, $(b)\diamondsuit$ $(AB)^* = B^* A^*$, (c) $(cA)^* = c A^*$ for $c \in \mathbb{R}$, $(d)\heartsuit$ $(A^*)^* = A$, $(e)\diamondsuit$ $(A^{-1})^* = (A^*)^{-1}$.

Solution: (a)
$$\begin{aligned}\langle \mathbf{x}, (A+B)^* \mathbf{y} \rangle_C &= \langle (A+B)\mathbf{x}, \mathbf{y} \rangle_K = \langle A\mathbf{x}, \mathbf{y} \rangle_K + \langle B\mathbf{x}, \mathbf{y} \rangle_K \\ &= \langle \mathbf{x}, A^*\mathbf{y} \rangle_C + \langle \mathbf{x}, B^*\mathbf{y} \rangle_C = \langle \mathbf{x}, (A^* + B^*)\mathbf{y} \rangle_C.\end{aligned}$$
Since this holds for all $\mathbf{x}, \mathbf{y}$, we conclude that $(A+B)^* = A^* + B^*$. ■

(d) $\langle (A^*)^*\mathbf{x}, \mathbf{y} \rangle_K = \langle \mathbf{x}, A^*\mathbf{y} \rangle_C = \langle A\mathbf{x}, \mathbf{y} \rangle_K$. ■

3.8. ♡ Let C, K be positive definite matrices defining inner products on $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Let A be an $m \times n$ matrix with adjoint A^*. Prove that $\mathbf{x}$ solves the inhomogeneous linear system $A\mathbf{x} = \mathbf{b}$ if and only if

$$\langle \mathbf{x}, A^*\mathbf{y} \rangle_C = \langle \mathbf{b}, \mathbf{y} \rangle_K \qquad \text{for all} \qquad \mathbf{y} \in \mathbb{R}^m. \tag{4.27}$$

Remark: Equation (4.27) is known as the *weak formulation* of the linear system. Its generalizations plays an essential role in the analysis of differential equations and their numerical approximations, [180, 192, 225].

Solution: By the definition of the adjoint, $\langle \mathbf{x}, A^*\mathbf{y} \rangle_C = \langle A\mathbf{x}, \mathbf{y} \rangle_K = \langle \mathbf{b}, \mathbf{y} \rangle_K$. Since this holds for all $\mathbf{y} \in \mathbb{R}^m$, it is equivalent to the system $A\mathbf{x} = \mathbf{b}$. ■

4.1. For each of the following matrices find bases (when they exist) for the

(*i*) image, (*ii*) coimage, (*iii*) kernel, and (*iv*) cokernel.

$$(a)\heartsuit \begin{pmatrix} 1 & -3 \\ 2 & -6 \end{pmatrix}, \quad (b)\heartsuit \begin{pmatrix} 1 & -3 & 1 \\ 0 & 2 & 0 \end{pmatrix}, \quad (c)\diamondsuit \begin{pmatrix} 0 & 0 & -8 \\ 1 & 2 & -1 \\ 2 & 4 & 6 \end{pmatrix}, \quad (d) \begin{pmatrix} 1 & 1 & 3 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 3 & 0 \end{pmatrix}.$$

Solution: (*a*) image: $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$; coimage: $\begin{pmatrix} 1 \\ -3 \end{pmatrix}$; kernel: $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$; cokernel: $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$.

(*b*) image: $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 2 \end{pmatrix}$; coimage: $\begin{pmatrix} 1 \\ -3 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$; kernel: $\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$; cokernel: $\{\mathbf{0}\}$ (no basis).

4.3. Find bases for the coimage and cokernel of the matrix $\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix}$ using (*a*) ♡ the dot product; (*b*) ♢ the weighted inner product $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + 2 v_2 w_2 + 3 v_3 w_3$. Make sure that the dimensions satisfy the formulas in the Fundamental Theorem 4.24.

Solution: (*a*) coimage: $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$; cokernel: $\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$.

4.5. ♡ *True or false*: nullity A = nullity A^*.

Solution: True if A is square; otherwise false.

5.1. Determine which of the following are orthogonal matrices: $(a)\heartsuit \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$,

$$(b) \begin{pmatrix} \frac{12}{13} & \frac{5}{13} \\ -\frac{5}{13} & \frac{12}{13} \end{pmatrix}, \quad (c)\heartsuit \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad (d) \begin{pmatrix} -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{pmatrix}, \quad (e)\diamondsuit \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \end{pmatrix}.$$

Solution: (*a*) Not orthogonal; (*c*) orthogonal.

5.4. *True or false*:
(a) ♡ A matrix whose columns form an orthogonal basis of $\mathbb{R}^n$ is an orthogonal matrix.
(b) ♢ A matrix whose rows form an orthonormal basis of $\mathbb{R}^n$ is an orthogonal matrix.
(c) An orthogonal matrix is symmetric if and only if it is a diagonal matrix.

Solution: (a) False — they must be an orthonormal basis.

5.5. Which of the indicated maps define isometries of the Euclidean plane?

$$(a)\heartsuit \begin{pmatrix} y \\ -x \end{pmatrix}, \quad (b)\diamondsuit \begin{pmatrix} x-y+1 \\ x+2 \end{pmatrix}, \quad (c)\heartsuit \frac{1}{\sqrt{2}}\begin{pmatrix} x+y-3 \\ x+y-2 \end{pmatrix}, \quad (d)\ \frac{1}{5}\begin{pmatrix} 3x+4y \\ -4x+3y+1 \end{pmatrix}.$$

Solution: (a) Isometry, (c) not an isometry.

5.6. Which of the following matrices are Euclidean norm-preserving?

$$(a)\heartsuit \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad (b)\heartsuit \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad (c)\diamondsuit \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}, \quad (d)\ \begin{pmatrix} \frac{2}{3} & \frac{1}{\sqrt{2}} \\ -\frac{1}{3} & 0 \\ -\frac{2}{3} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Solution: (a) Not norm-preserving, because it maps from $\mathbb{R}^3$ to $\mathbb{R}^2$; (b) norm-preserving.

6.1. Using the dot product, write out the projection matrix corresponding to the subspaces spanned by

$$(a)\heartsuit \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \quad (b)\ \begin{pmatrix} \frac{2}{3} \\ -\frac{2}{3} \\ \frac{1}{3} \end{pmatrix}, \quad (c)\heartsuit \begin{pmatrix} \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \quad (d)\diamondsuit \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}.$$

Solution: (a) $\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$; (c) $\begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$.

6.2. Write out the projection matrices onto the orthogonal complements of the subspaces in Exercise 6.1.

Solution: (a) $\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$; (c) $\begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$.

6.3. ♡ Prove that a projection matrix is positive semidefinite. When is it positive definite?

Solution: Since $P = U U^T$ where U is a $k \times n$ matrix with orthonormal columns, it is the Gram matrix corresponding to U^T, and hence positive semidefinite. It is positive definite if and only if $\operatorname{rank} U = \operatorname{rank} U^T = n$ which implies U has size $n \times n$ and hence is an orthogonal matrix. Thus, $P = U U^T = \mathrm{I}$ is the only positive definite projection matrix. ■

7.1. Find the QR factorization of the following matrices: $(a)\heartsuit \begin{pmatrix} 1 & -3 \\ 2 & 1 \end{pmatrix}$, $(b) \begin{pmatrix} 4 & 3 \\ 3 & 2 \end{pmatrix}$,

$(c)\heartsuit \begin{pmatrix} 2 & 1 & -1 \\ 0 & 1 & 3 \\ -1 & -1 & 1 \end{pmatrix}$, $(d) \begin{pmatrix} 0 & 1 & 2 \\ -1 & 1 & 1 \\ -1 & 1 & 3 \end{pmatrix}$, $(e)\diamondsuit \begin{pmatrix} 0 & 0 & 2 \\ 0 & 4 & 1 \\ -1 & 0 & 1 \end{pmatrix}$, $(f) \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 2 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$.

Solution: (a) $\begin{pmatrix} 1 & -3 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt5} & -\frac{2}{\sqrt5} \\ \frac{2}{\sqrt5} & \frac{1}{\sqrt5} \end{pmatrix} \begin{pmatrix} \sqrt5 & -\frac{1}{\sqrt5} \\ 0 & \frac{7}{\sqrt5} \end{pmatrix}$,

(c) $\begin{pmatrix} 2 & 1 & -1 \\ 0 & 1 & 3 \\ -1 & -1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt5} & -\frac{1}{\sqrt{30}} & \frac{1}{\sqrt6} \\ 0 & \sqrt{\frac56} & \frac{1}{\sqrt6} \\ -\frac{1}{\sqrt5} & -\sqrt{\frac{2}{15}} & \sqrt{\frac23} \end{pmatrix} \begin{pmatrix} \sqrt5 & \frac{3}{\sqrt5} & -\frac{3}{\sqrt5} \\ 0 & \sqrt{\frac65} & 7\sqrt{\frac{2}{15}} \\ 0 & 0 & 2\sqrt{\frac23} \end{pmatrix}$.

7.2. For each of the following linear systems, find the QR factorization of the coefficient matrix, and then use your factorization to solve the system: $(a)\heartsuit \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$,

$(b)\diamondsuit \begin{pmatrix} 2 & 1 & -1 \\ 1 & 0 & 2 \\ 2 & -1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}$, $(c) \begin{pmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$.

Solution: (a) $\begin{pmatrix} \frac{1}{\sqrt2} & \frac{1}{\sqrt2} \\ -\frac{1}{\sqrt2} & \frac{1}{\sqrt2} \end{pmatrix} \begin{pmatrix} \sqrt2 & -\frac{1}{\sqrt2} \\ 0 & \frac{5}{\sqrt2} \end{pmatrix}$, $\quad \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -\frac75 \\ \frac15 \end{pmatrix}$.

7.3. Determine the rank of the following matrices using the extended QR method:

$(a)\heartsuit \begin{pmatrix} 1 & 2 \\ -2 & -4 \end{pmatrix}$, $(b) \begin{pmatrix} 3 & -1 & -2 \\ -6 & 2 & 1 \end{pmatrix}$, $(c)\heartsuit \begin{pmatrix} 2 & -5 & -1 \\ 1 & -6 & -4 \\ 3 & -4 & 2 \end{pmatrix}$, $(d)\diamondsuit \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 6 & 4 & 7 \\ 1 & 2 & 2 & 3 \\ 3 & 6 & 5 & 8 \end{pmatrix}$.

Solution: (a) $\begin{pmatrix} 1 & 2 \\ -2 & -4 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt5} \\ -\frac{2}{\sqrt5} \end{pmatrix} \begin{pmatrix} \sqrt5 & 2\sqrt5 \end{pmatrix}$, rank $= 1$;

(c) $\begin{pmatrix} 2 & -5 & -1 \\ 1 & -6 & -4 \\ 3 & -4 & 2 \end{pmatrix} = \begin{pmatrix} \frac{4}{\sqrt{14}} & -\frac{1}{\sqrt{21}} \\ \frac{1}{\sqrt{14}} & -\frac{4}{\sqrt{21}} \\ \frac{3}{\sqrt{14}} & \frac{2}{\sqrt{21}} \end{pmatrix} \begin{pmatrix} \sqrt{14} & -2\sqrt{14} & 0 \\ 0 & \sqrt{21} & \sqrt{21} \end{pmatrix}$, rank $= 2$.

7.4. Use the QR method to compute the least squares solution to the linear system $A\mathbf{x} = \mathbf{b}$ when (a)♡ $A = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$; (b)♢ $A = \begin{pmatrix} 1 & 0 \\ 2 & -1 \\ 3 & 2 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$;

(c)♡ $A = \begin{pmatrix} 2 & 1 & -1 \\ 1 & -2 & 0 \\ 3 & -1 & -1 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$; (d) $A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & -2 & -1 \\ 1 & 0 & -1 \\ 5 & 0 & 1 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$.

Solution: (a) $Q = \begin{pmatrix} \frac{1}{\sqrt{6}} \\ -\frac{4}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix}$, $R = \begin{pmatrix} \sqrt{6} \end{pmatrix}$; solve $Rx = Q^T\mathbf{b} = \sqrt{\frac{3}{2}}$ for $x = \frac{1}{2}$.

(c) $Q = \begin{pmatrix} \frac{4}{\sqrt{14}} & -\frac{4}{\sqrt{42}} \\ \frac{1}{\sqrt{14}} & -\frac{5}{\sqrt{42}} \\ \frac{3}{\sqrt{14}} & -\frac{1}{\sqrt{42}} \end{pmatrix}$, $R = \begin{pmatrix} \sqrt{14} & -\frac{3}{\sqrt{14}} & -\frac{5}{\sqrt{14}} \\ 0 & \frac{15}{\sqrt{42}} & -\frac{3}{\sqrt{42}} \end{pmatrix}$;

solve $R\mathbf{x} = Q^T\mathbf{b} = \begin{pmatrix} -\frac{1}{\sqrt{14}} \\ \frac{5}{\sqrt{42}} \end{pmatrix}$ for $\mathbf{x} = \begin{pmatrix} \frac{2}{5}z \\ \frac{1}{3} + \frac{1}{5}z \\ z \end{pmatrix}$, where z is a free variable, in accordance with nullity $A = 1$.

Solution: The missing code in the Python notebook from this section is below.

```
R[0,0] = np.linalg.norm(A[:,0])
Q[:,0] = A[:,0] / R[0,0]
for k in range(1,m):
    #First orthogonalization
    s = Q[:,:k].T@A[:,k]
    v = A[:,k] - Q[:,:k]@s
    #Re-orthogonalization
    t = Q[:,:k].T@v
    xk = v - Q[:,:k]@t
```

8.1. Compute (i) the 1 matrix norm, (ii) the ∞ matrix norm, and (iii) the Frobenius norm of the following matrices:

(a)♡ $\begin{pmatrix} 2 & -2 \\ -3 & 5 \end{pmatrix}$, (b)♢ $\begin{pmatrix} \frac{5}{3} & \frac{4}{3} \\ -\frac{7}{6} & -\frac{5}{6} \end{pmatrix}$, (c)♡ $\begin{pmatrix} 1 & -2 & 3 \\ -1 & 0 & 1 \\ 2 & -1 & 1 \end{pmatrix}$, (d) $\begin{pmatrix} 0 & .2 & .8 \\ -.3 & 0 & .1 \\ -.4 & .1 & 0 \end{pmatrix}$.

Solution: (a) (i) 7, (ii) 8, (iii) $\sqrt{42} \approx 6.4807$ (c) (i) 5, (ii) 6, (iii) $\sqrt{22} \approx 4.6904$.

8.2. Let $A = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}$. Compute the natural matrix norm $\| A \|$ using the following norms on $\mathbb{R}^2$: (a)♡ the 1 norm; (b)♢ the ∞ norm; (c)♡ the weighted 1 norm $\| \mathbf{v} \| = 2 | v_1 | + 3 | v_2 |$; (d) the weighted ∞ norm $\| \mathbf{v} \| = \max \{ 2 | v_1 |, 3 | v_2 | \}$.

Solution: (a) 3, (c) $\| A \| = \frac{8}{3}$. The unit sphere for this norm is the diamond with corners $\pm \left(\frac{1}{2}, 0 \right)^T$, $\pm \left(0, \frac{1}{3} \right)^T$. It is mapped to the parallelogram with corners $\pm \left(\frac{1}{2}, \frac{1}{2} \right)^T$, $\pm \left(\frac{1}{3}, -\frac{2}{3} \right)^T$, with respective norms $\frac{5}{2}$ and $\frac{8}{3}$, and so $\| A \| = \max\{ \| A \mathbf{v} \| \mid \| \mathbf{v} \| = 1 \} = \frac{8}{3}$.

8.3.♡ Find a matrix A such that $\| A^2 \|_\infty \neq \| A \|_\infty^2$.

Solution: Example: $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ has $\| A \|_\infty = 2$, while $A^2 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ has $\| A^2 \|_\infty = 3$.

8.7.♡ *True or false*: If $\| A \|_F < 1$, then $A^k \to \mathrm{O}$ as $k \to \infty$.

Solution: True, since by (4.88), $\| A^k \|_F \le \| A \|_F^k \to 0$ as $k \to \infty$.

Chapter 5

Eigenvalues and Singular Values

1.1. Find the eigenvalues and eigenvectors of the following 2×2 matrices:

$$(a) \heartsuit \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad (b) \diamondsuit \begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix}, \qquad (c) \begin{pmatrix} 1 & -\frac{2}{3} \\ \frac{1}{2} & -\frac{1}{6} \end{pmatrix}, \qquad (d) \heartsuit \begin{pmatrix} 3 & 1 \\ -1 & 1 \end{pmatrix}.$$

Solution: (a) Eigenvalues: $1, -1$; eigenvectors: $\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

(d) Eigenvalue: 2; eigenvector: $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

1.2. Write down (a) $\heartsuit$ a 2×2 matrix that has 0 as one of its eigenvalues and $(1, 2)^T$ as a corresponding eigenvector; (b) a 3×3 matrix that has $(1, 2, 3)^T$ as an eigenvector for the eigenvalue -1; (c) $\heartsuit$ a 4×4 matrix that has -4 as an eigenvalue with multiplicity 2.

Solution: Examples: (a) $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, (c) $\begin{pmatrix} -4 & 0 & 0 & 0 \\ 0 & -4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.

1.3. Find all eigenvalues and eigenvectors of (a) $\heartsuit$ the $n \times n$ zero matrix O; (b) the $n \times n$ identity matrix I; (c) $\diamondsuit$ the $n \times n$ matrix $E = \mathbf{1}\mathbf{1}^T$ with every entry equal to 1..

Solution: (a) Since $\mathrm{O}\mathbf{v} = \mathbf{0} = 0\mathbf{v}$, we conclude that 0 is the only eigenvalue; all nonzero vectors $\mathbf{v} \neq \mathbf{0}$ are eigenvectors.

1.6. Let A be a square matrix. (a) $\heartsuit$ Explain in detail why every nonzero scalar multiple of an eigenvector of A is also an eigenvector. (b) $\heartsuit$ Show that every nonzero linear combination of two eigenvectors $\mathbf{v}, \mathbf{w}$ corresponding to the *same* eigenvalue is also an eigenvector. (c) $\diamondsuit$ Prove that a linear combination $c\mathbf{v} + d\mathbf{w}$, with $c, d \neq 0$, of two eigenvectors corresponding to *different* eigenvalues is never an eigenvector.

Solution: (a) If $A\mathbf{v} = \lambda\mathbf{v}$, then $A(c\mathbf{v}) = cA\mathbf{v} = c\lambda\mathbf{v} = \lambda(c\mathbf{v})$ and so $c\mathbf{v}$ satisfies the eigenvector equation for the eigenvalue λ. Moreover, since $\mathbf{v} \neq \mathbf{0}$, also $c\mathbf{v} \neq \mathbf{0}$ for $c \neq 0$, and so $c\mathbf{v}$ is a bona fide eigenvector. ■

(b) If $A\mathbf{v} = \lambda\mathbf{v}$, $A\mathbf{w} = \lambda\mathbf{w}$, then $A(c\mathbf{v}+d\mathbf{w}) = cA\mathbf{v}+dA\mathbf{w} = c\lambda\mathbf{v}+d\lambda\mathbf{w} = \lambda(c\mathbf{v}+d\mathbf{w})$. Since $\mathbf{v}.\mathbf{w}$ are linearly independent, we have $c\mathbf{v}+d\mathbf{w} \neq \mathbf{0}$, and hence it is an eigenvector for the eigenvalue λ. ■

1.8. ♡ (a) Show that if λ is an eigenvalue of A, then λ^2 is an eigenvalue of A^2. (b) Is the converse valid: if μ is an eigenvalue of A^2, then $\sqrt{\mu}$ is an eigenvalue of A?

Solution: (a) If $A\mathbf{v} = \lambda\mathbf{v}$ then $A^2\mathbf{v} = \lambda A\mathbf{v} = \lambda^2\mathbf{v}$, and hence $\mathbf{v}$ is also an eigenvector of A^2 with eigenvalue λ^2. ■ (b) Not necessarily; possibly $-\sqrt{\mu}$ is the only eigenvalue.

1.11. *True or false*:

(a) ♡ If λ is an eigenvalue of both A and B, then it is an eigenvalue of the sum $A+B$.
(b) ♢ If $\mathbf{v}$ is an eigenvector of both A and B, then it is an eigenvector of $A+B$.
(c) If λ is an eigenvalue of A and μ is an eigenvalue of B, then $\lambda\mu$ is an eigenvalue of the matrix product $C = AB$.

Solution: (a) False. For example, 0 is an eigenvalue of both $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}$, but the eigenvalues of $A+B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ are ± 1.

2.1. Which of the following matrices are complete? For those that are, exhibit an eigenvector basis of $\mathbb{R}^2$. For those that are not, what is the dimension of the subspace of $\mathbb{R}^2$ spanned by the eigenvectors? (a) ♡ $\begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$, ($b$) ♢ $\begin{pmatrix} 1 & 3 \\ -3 & 1 \end{pmatrix}$, ($c$) ♡ $\begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}$, ($d$) $\begin{pmatrix} 3 & 3 \\ 1 & 5 \end{pmatrix}$.

Solution: (a) Complete: eigenvalues: $-2, 4$; the eigenvectors $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ form a basis.

(c) Not complete: the only eigenvalue is 1, and there is only one independent eigenvector $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ spanning a one-dimensional subspace of $\mathbb{R}^2$.

2.2. Find the spectral radius of the following matrices. Which are convergent?

(a) ♡ $\begin{pmatrix} .3 & -.4 \\ -.2 & .6 \end{pmatrix}$, ($b$) ♢ $\begin{pmatrix} 0 & \frac{4}{5} \\ \frac{3}{5} & \frac{2}{3} \end{pmatrix}$, ($c$) ♡ $\begin{pmatrix} .3 & 2.2 & -1.7 \\ 0 & -.6 & 4.1 \\ 0 & 0 & .5 \end{pmatrix}$, ($d$) $\begin{pmatrix} 0 & 1 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix}$.

Solution: (a) 0.7702, convergent (c) 0.6, convergent.

2.3. Use (5.19) to write down an explicit formula for the k-th power of the following matrices:

(a) ♡ $\begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}$, ($b$) $\begin{pmatrix} 4 & 1 \\ -2 & 1 \end{pmatrix}$, ($c$) ♡ $\begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{pmatrix}$, ($d$) ♢ $\begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 1 \end{pmatrix}$.

Solution: (a) $\begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 6^k & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{2}{5} & \frac{1}{5} \\ -\frac{1}{5} & \frac{2}{5} \end{pmatrix}$,

(c) $\begin{pmatrix} -1 & 1 & 1 \\ -2 & 0 & 1 \\ 2 & 0 & 0 \end{pmatrix} \begin{pmatrix} (-1)^k & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & \frac{1}{2} \\ 1 & -1 & -\frac{1}{2} \\ 0 & 1 & 1 \end{pmatrix}$,

which is $\begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{pmatrix}$ if k is odd and $\begin{pmatrix} 1 & -1 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{pmatrix}$ if $k > 0$ is even.

2.5. ♡ Prove that if A is a complete matrix, then so is $cA + d\,\mathrm{I}$, where c, d are any scalars.

Solution: According to Exercise 1.7, every eigenvector of A is an eigenvector of $cA + d\,\mathrm{I}$ with eigenvalue $c\lambda + d$, and hence if A has an eigenvector basis, so does $cA + d\,\mathrm{I}$. ■

2.9. *True or false*: (a) ♡ $\rho(cA) = c\,\rho(A)$, (b) $\rho(V^{-1}AV) = \rho(A)$, (c) ♡ $\rho(A^2) = \rho(A)^2$, (d) ♡ $\rho(A^{-1}) = 1/\rho(A)$, (e) ♢ $\rho(A+B) = \rho(A) + \rho(B)$, (f) $\rho(AB) = \rho(A)\,\rho(B)$.

Solution: (a) False: $\rho(cA) = |\,c\,|\,\rho(A)$.
(c) True, since the eigenvalues of A^2 are the squares of the eigenvalues of A.
(d) False, e.g., if $A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ then $\rho(A) = 2$, but $A^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix}$ and $\rho(A) = 1$.

3.1. Find the eigenvalues and an orthonormal eigenvector basis for the following symmetric matrices, and then write out their spectral factorization. Use this to determine which are positive definite. (a) ♡ $\begin{pmatrix} 2 & 6 \\ 6 & -7 \end{pmatrix}$ (b) $\begin{pmatrix} 5 & -2 \\ -2 & 5 \end{pmatrix}$, (c) ♢ $\begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & 3 \\ 4 & 3 & 1 \end{pmatrix}$.

Solution: (a) Eigenvalues: $5, -10$; eigenvectors: $\dfrac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \dfrac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ -2 \end{pmatrix}$,

spectral factorization: $\begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{pmatrix}\begin{pmatrix} 5 & 0 \\ 0 & -10 \end{pmatrix}\begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{pmatrix}$.

3.2. Construct a symmetric matrix that has the following eigenvalues and associated eigenvectors, or explain why none exists:
(a) ♡ $\lambda_1 = -2$, $\mathbf{v}_1 = (\,1, -1\,)^T$, $\lambda_2 = 1$, $\mathbf{v}_2 = (\,1, 1\,)^T$, (b) ♢ $\lambda_1 = 3$, $\mathbf{v}_1 = (\,2, -1\,)^T$, $\lambda_2 = -1$, $\mathbf{v}_2 = (\,-1, 2\,)^T$, (c) $\lambda_1 = 2$, $\mathbf{v}_1 = (\,2, 1\,)^T$, $\lambda_2 = 2$, $\mathbf{v}_2 = (\,1, 2\,)^T$.

Solution: (a) $\begin{pmatrix} -\frac{1}{2} & \frac{3}{2} \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}$.

3.5. Prove the properties (b) ♡ , (c) ♢ , (d) listed on page 140.

Solution: (b) According to (4.24), $\widetilde{S}$ is self-adjoint with respect to the inner product defined by C provided $\widetilde{S}^T C = C\,\widetilde{S}$. Thus $S^T = C^{-1/2}\,\widetilde{S}^T C^{1/2} = C^{1/2}\,\widetilde{S}\,C^{-1/2} = S$. Moreover, if $\mathbf{x} = C^{1/2}\mathbf{y}$, then $\mathbf{x} \neq \mathbf{0}$ if and only if $\mathbf{y} \neq \mathbf{0}$, hence $\mathbf{y}^T S\,\mathbf{y} = \mathbf{x}^T \widetilde{S}\,\mathbf{x} > 0$ for all $\mathbf{y} \neq \mathbf{0}$. ■

3.8. ♡ Given an inner product on $\mathbb{R}^n$, let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be an orthonormal basis. Prove that they form an eigenvector basis for some self-adjoint $n \times n$ matrix S. Can you characterize all such matrices? Under what conditions can you construct such an S that is positive definite?

Solution: The simplest is $S = \mathrm{I}$. More generally, any matrix of the form $S = U^*\Lambda U$, where $U = (\,\mathbf{u}_1\ \mathbf{u}_2\ \ldots\ \mathbf{u}_n\,)$ and Λ is any real diagonal matrix. The matrix is positive definite if and only if Λ has all positive diagonal entries. ■

3.9. ♡ Find a non-symmetric 2×2 matrix S with real eigenvalues that does not satisfy the inequalities (5.62).

Solution:
For example, $S = \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix}$ has only the eigenvalue $\lambda_1 = 1$, but $\widetilde{S} = (2)$ has eigenvalue 2.

3.11. ♡ Write down two self adjoint positive definite matrices whose Hadamard product is not positive definite.

Solution: Let $C = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$, so $C^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix}$. Then $R = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} \\ 1 & 1 \end{pmatrix}$ and $S = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -\frac{1}{2} \\ -1 & 1 \end{pmatrix}$ are both self adjoint and positive definite, but their Hadamard product $R \circ S = \begin{pmatrix} 1 & -\frac{1}{4} \\ -1 & 1 \end{pmatrix}$ is not self adjoint for the C inner product.

3.12. Compute the generalized eigenvalues and eigenvectors for the following matrix pairs. Verify orthogonality of the eigenvectors under the appropriate inner product.

(a) ♡ $K = \begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}, \quad C = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$; (b) $K = \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$;

(c) $K = \begin{pmatrix} 2 & -1 \\ -1 & 4 \end{pmatrix}, \ C = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$; (d) ♢ $K = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 8 & 2 \\ 0 & 2 & 1 \end{pmatrix}, \ C = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 1 \end{pmatrix}$.

Solution: (a) Eigenvalues: $\frac{5}{3}, \frac{1}{2}$; eigenvectors: $\begin{pmatrix} -3 \\ 1 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix}$.

4.2. Write down and solve optimization principles characterizing the largest and the smallest eigenvalues of the following positive definite matrices:

(a) ♡ $\begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix}$, (b) ♢ $\begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}$, (c) ♡ $\begin{pmatrix} 3 & 0 & -1 \\ 0 & 3 & 0 \\ -1 & 0 & 3 \end{pmatrix}$, (d) $\begin{pmatrix} 4 & -1 & -2 \\ -1 & 4 & -1 \\ -2 & -1 & 4 \end{pmatrix}$.

Solution: (a)
$$\frac{5+\sqrt{5}}{2} = \max\left\{ 2x^2 - 2xy + 3y^2 \;\middle|\; x^2 + y^2 = 1 \right\},$$
$$\frac{5-\sqrt{5}}{2} = \min\left\{ 2x^2 - 2xy + 3y^2 \;\middle|\; x^2 + y^2 = 1 \right\};$$
(c)
$$4 = \max\left\{ 3x^2 - 2xz + 3y^2 + 3z^2 \;\middle|\; x^2 + y^2 + z^2 = 1 \right\},$$
$$2 = \min\left\{ 3x^2 - 2xz + 3y^2 + 3z^2 \;\middle|\; x^2 + y^2 + z^2 = 1 \right\}.$$

4.3. Write down and solve a maximization principle that characterizes the middle eigenvalue of the matrices in parts (c) and (d) of Exercise 4.2.

Solution: (c) $3 = \max\left\{ 3x^2 - 2xz + 3y^2 + 3z^2 \;\middle|\; x^2 + y^2 + z^2 = 1, x - z = 0 \right\}$.

4.6. Write out optimization principles for the largest and smallest generalized eigenvalues of the matrix pairs in Exercise 3.12.

Solution: (a) $\frac{5}{3} = \max\left\{\, 3x^2 - 2xy + 2y^2 \;\middle|\; 2x^2 + 3y^2 = 1 \,\right\}$,
$\frac{1}{2} = \min\left\{\, 3x^2 - 2xy + 2y^2 \;\middle|\; 2x^2 + 3y^2 = 1 \,\right\}$.

5.1. Find the explicit formula for the solution to the following linear iterative systems:
(a) ♡ $x_{k+1} = x_k - 2y_k,\ \ y_{k+1} = -2x_k + y_k,\ \ x_0 = 1,\ \ y_0 = 0;$
(b) ♢ $x_{k+1} = x_k - \frac{2}{3}y_k,\ \ y_{k+1} = \frac{1}{2}x_k - \frac{1}{6}y_k,\ \ x_0 = -2,\ \ y_0 = 3;$
(c) $x_{k+1} = x_k - y_k,\ \ y_{k+1} = -x_k + 5y_k,\ \ x_0 = 1,\ \ y_0 = 0.$

Solution: (a) $x_k = \dfrac{3^k + (-1)^k}{2}, \quad y_k = \dfrac{-3^k + (-1)^k}{2}.$

5.2. Use your answers from Exercise 2.3 to solve the following iterative systems:
(a) ♡ $x_{k+1} = 5x_k + 2y_k,\ \ y_{k+1} = 2x_k + 2y_k,\ \ x_0 = 1,\ \ y_0 = -1;$
(b) $x_{k+1} = 4x_k + y_k,\ \ y_{k+1} = -2x_k + y_k,\ \ x_0 = 1,\ \ y_0 = -1;$
(c) ♡ $x_{k+1} = x_k - y_k,\ \ y_{k+1} = z_k,\ \ z_{k+1} = -z_k,\ \ \mathbf{x}_0 = 1,\ \ y_0 = 3,\ \ z_0 = 2;$
(d) ♢ $x_{k+1} = x_k + y_k + 2z_k,\ \ y_{k+1} = x_k + 2y_k + z_k,\ \ z_{k+1} = 2x_k + y_k + z_k,$
$\mathbf{x}_0 = 1,\ \ y_0 = 0,\ \ z_0 = 1.$

Solution: (a) $\begin{pmatrix} x_k \\ y_k \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 6^k/5 \\ -\frac{3}{5} \end{pmatrix}$; (c) $\begin{pmatrix} x_k \\ y_k \\ z_k \end{pmatrix} = \begin{pmatrix} -1 & 1 & 1 \\ -2 & 0 & 1 \\ 2 & 0 & 0 \end{pmatrix} \begin{pmatrix} (-1)^k \\ -3 \\ 0 \end{pmatrix}$,
for $k \geq 1$, which equals $\begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix}$ if k is odd and $\begin{pmatrix} -4 \\ -2 \\ 2 \end{pmatrix}$ if $k > 0$ is even.

5.3. ♡ Explain why the j-th column $\mathbf{c}_j^{(k)}$ of the matrix power A^k satisfies the linear iterative system $\mathbf{c}_j^{(k+1)} = A\,\mathbf{c}_j^{(k)}$ with initial data $\mathbf{c}_j^{(0)} = \mathbf{e}_j$, the j-th standard basis vector.

Solution: Since matrix multiplication acts column-wise, as per formula (3.8), the j-th column of the matrix equation $A^{k+1} = A\,A^k$ is $\mathbf{c}_j^{(k+1)} = A\,\mathbf{c}_j^{(k)}$. Moreover, $A^0 = \mathrm{I}$ has j-th column $\mathbf{c}_j^{(0)} = \mathbf{e}_j$. ■

5.6. *True or false*: (a) ♡ If A is convergent, then A^2 is convergent.
(b) ♢ If A is convergent, then $A^T A$ is convergent.

Solution: (a) True by part (c) of Exercise 2.9.

5.8. Determine if the following matrices are regular transition matrices. If so, find the associated probability eigenvector.

(a) ♡ $\begin{pmatrix} \frac{1}{2} & \frac{1}{3} \\ \frac{3}{4} & \frac{2}{3} \end{pmatrix}$, (b) ♡ $\begin{pmatrix} \frac{1}{4} & \frac{2}{3} \\ \frac{3}{4} & \frac{1}{3} \end{pmatrix}$, (c) $\begin{pmatrix} 0 & \frac{1}{5} \\ 1 & \frac{4}{5} \end{pmatrix}$, (d) ♢ $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, (e) $\begin{pmatrix} .3 & .5 & .2 \\ .3 & .2 & .5 \\ .4 & .3 & .3 \end{pmatrix}$.

Solution: (a) Not a transition matrix; (b) regular transition matrix: $\left(\frac{8}{17}, \frac{9}{17}\right)^T$.

5.9. ♡ Explain why the irregular Markov process with transition matrix $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ does not reach a steady state.

Solution: If $\mathbf{x}_0 = (a, b)^T$ is the initial state vector, then the subsequent state vectors switch back and forth between $(b, a)^T$ and $(a, b)^T$. At each step in the process, all of the population in state 1 goes to state 2 and vice versa, so the system never settles down.

5.10. ♡ A certain plant species has either red, pink, or white flowers, depending on its genotype. If you cross a pink plant with any other plant, the probability distribution of the offspring is prescribed by the transition matrix $A = \begin{pmatrix} .5 & .25 & 0 \\ .5 & .5 & .5 \\ 0 & .25 & .5 \end{pmatrix}$. On average, if you continue crossing with only pink plants, what percentage of the three types of flowers would you expect to see in your garden?

Solution: 25% red, 50% pink, 25% pink.

6.1. Use the power method to approximate the dominant eigenvalue and associated eigenvector of the following matrices. Write your code in Python and compare to the output of `numpy.linalg.eig`.

(a) ♡ $\begin{pmatrix} -1 & -2 \\ 3 & 4 \end{pmatrix}$, (b) ♢ $\begin{pmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 3 \end{pmatrix}$, (c) ♡ $\begin{pmatrix} -2 & 0 & 1 \\ -3 & -2 & 0 \\ -2 & 5 & 4 \end{pmatrix}$, (d) $\begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}$.

Solution: In all cases, we use the normalized version (5.86) starting with $\mathbf{x}_0 = \mathbf{e}_1$; the answers are correct to 4 decimal places.

(a) After 17 iterations, $\lambda = 2.00002$, $\mathbf{u} = (-.55470, .83205)^T$.

(c) After 121 iterations, $\lambda = -3.30282$, $\mathbf{u} = (.35356, .81416, -.46059)^T$.

6.3. ♡ Prove that, for the normalized iterative method (5.86), $\| A\mathbf{y}_k \| \to |\lambda_1|$. Assuming λ_1 is real, explain how to deduce its sign.

Solution: Since $\mathbf{x}_k \to \lambda_1^k c_1 \mathbf{v}_1$ as $k \to \infty$, we have

$$\mathbf{y}_k = \frac{\mathbf{x}_k}{\|\mathbf{x}_k\|} \longrightarrow \frac{c_1 \lambda_1^k \mathbf{u}_1}{|c_1|\,|\lambda_1|^k \|\mathbf{v}_1\|} = \begin{cases} \mathbf{u}_1, & \lambda_1 > 0, \\ (-1)^k \mathbf{u}_1, & \lambda_1 < 0, \end{cases} \quad \text{where} \quad \mathbf{u}_1 = (\operatorname{sign} c_1) \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$$

is one of the two real unit eigenvectors. Moreover, $A\mathbf{y}_k \to \begin{cases} \lambda_1 \mathbf{u}_1, & \lambda_1 > 0, \\ (-1)^k \lambda_1 \mathbf{u}_1, & \lambda_1 < 0, \end{cases}$ so $\| A\mathbf{y}_k \| \to |\lambda_1|$. If $\lambda_1 > 0$, the iterates $\mathbf{y}_k \to \mathbf{u}_1$ converge to one of the two dominant unit eigenvectors, whereas if $\lambda_1 < 0$, the iterates $\mathbf{y}_k \to (-1)^k \mathbf{u}_1$ switch back and forth between the two unit eigenvectors. ■

6.5. ♡ *The Inverse Power Method.* Let A be a nonsingular matrix. (*i*) Show that the eigenvalues of A^{-1} are the reciprocals $1/\lambda$ of the eigenvalues of A. How are the eigenvectors related? (*ii*) Show how to use the power method on A^{-1} to produce the smallest (in modulus) eigenvalue of A. (*iii*) What is the rate of convergence of the algorithm? (*iv*) Design a practical iterative algorithm based on the QR decomposition of A. (*v*) Apply your algorithm to find the smallest eigenvalues and associated eigenvectors of the matrices in Exercise 6.1.

Solution: (*i*) First note that 0 cannot be an eigenvalue because A is nonsingular. If $A\mathbf{v} = \lambda\mathbf{v}$ then $A^{-1}\mathbf{v} = \lambda^{-1}\mathbf{v}$, and so $\mathbf{v}$ is also the eigenvector of A^{-1}.

(*ii*) If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, and we assume that $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| > 0$ (otherwise, see Exercise 6.4), then $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$ are the eigenvalues of A^{-1}, and $1/|\lambda_n| > 1/|\lambda_{n-1}| > \cdots > 1/|\lambda_1|$ and so $1/\lambda_n$ is the dominant eigenvalue of A^{-1}. Thus, applying the power method to A^{-1} will produce the reciprocal of the smallest (in absolute value) eigenvalue of A and its corresponding eigenvector.

(*iii*) The rate of convergence of the algorithm is the ratio $|\lambda_n/\lambda_{n-1}|$ of the moduli of the smallest two eigenvalues.

(*iv*) Once we factor $A = QR$, we can solve the iteration equation $A\mathbf{u}^{(k+1)} = \mathbf{x}_k$ by rewriting it in the form $R\mathbf{x}_{k+1} = Q^T\mathbf{x}_k$, and then using Back Substitution to solve for $\mathbf{x}_{k+1}$. As we know, this is much faster than computing A^{-1}.

(*a*) After 15 iterations, we obtain $\lambda = .99998$, $\mathbf{u} = (.70711, -.70710)^T$.

(*c*) After 6 iterations, we obtain $\lambda = .30277$, $\mathbf{u} = (.35355, -.46060, .81415)^T$.

6.10. Apply orthogonal iteration to the following symmetric matrices to find their eigenvalues and eigenvectors to 2 decimal places:

$$(a)\heartsuit \begin{pmatrix} 1 & 2 \\ 2 & 6 \end{pmatrix}, \quad (b) \begin{pmatrix} 3 & -1 \\ -1 & 5 \end{pmatrix}, \quad (c)\heartsuit \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 3 \\ 0 & 3 & 1 \end{pmatrix}, \quad (d)\diamondsuit \begin{pmatrix} 2 & 5 & 0 \\ 5 & 0 & -3 \\ 0 & -3 & 3 \end{pmatrix},$$

$$(e)\diamondsuit \begin{pmatrix} 3 & -1 & 0 & 0 \\ -1 & 3 & -1 & 0 \\ 0 & -1 & 3 & -1 \\ 0 & 0 & -1 & 3 \end{pmatrix}, \quad (f) \begin{pmatrix} 6 & 1 & -1 & 0 \\ 1 & 8 & 1 & -1 \\ -1 & 1 & 4 & 1 \\ 0 & -1 & 1 & 3 \end{pmatrix}.$$

Solution: (*a*) Eigenvalues: 6.7016, .2984; eigenvectors: $\begin{pmatrix} .3310 \\ .9436 \end{pmatrix}, \begin{pmatrix} .9436 \\ -.3310 \end{pmatrix}$.

(*c*) Eigenvalues: 4.7577, 1.9009, −1.6586; eigenvectors: $\begin{pmatrix} .2726 \\ .7519 \\ .6003 \end{pmatrix}, \begin{pmatrix} .9454 \\ -.0937 \\ -.3120 \end{pmatrix}, \begin{pmatrix} -.1784 \\ .6526 \\ -.7364 \end{pmatrix}$.

6.13.♡ Show that applying orthogonal iteration to the matrix $A = \begin{pmatrix} 4 & -1 & 1 \\ -1 & 7 & 2 \\ 1 & 2 & 7 \end{pmatrix}$, starting with the initial matrix $S_0 = \mathrm{I}$, eventually results in a diagonal matrix with the eigenvalues on the diagonal, but not in decreasing order. Explain why. Try changing the initial condition S_0; does that produce the eigenvalues in the correct order?

Solution: The iterates converge to the diagonal matrix $\begin{pmatrix} 6 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 3 \end{pmatrix}$. The eigenvalues appear on along the diagonal, but not in decreasing order, because, when the eigenvalues are listed in decreasing order, the corresponding eigenvector matrix $S = \begin{pmatrix} 0 & 1 & 1 \\ 1 & -1 & 1 \\ -2 & -1 & 1 \end{pmatrix}$ does not satisfy the regularity condition (5.94), and so Theorem 5.71 does not apply.

6.14.♡ Assume that orthogonal iteration applied to a symmetric positive semidefinite matrix A converges to an $n \times k$ matrix Q, whose columns are orthonormal, and a $k \times k$ upper triangular matrix R, whose diagonal entries are positive. Then Q and R satisfy $AQ = QR$. Show that the columns of Q are eigenvectors of A, and R is a diagonal matrix containing the corresponding eigenvalues.

Solution: We start with the first two columns of $AQ = QR$, which read

$$A\,\mathbf{q}_1 = r_{11}\,\mathbf{q}_1, \qquad A\,\mathbf{q}_2 = r_{12}\,\mathbf{q}_1 + r_{22}\,\mathbf{q}_2.$$

Thus, $\mathbf{q}_1$ is an eigenvector with eigenvalue r_{11}. Taking the dot product of the second equation with $\mathbf{q}_1$ and using orthogonality produces

$$r_{12} = \mathbf{q}_1 \cdot (r_{12}\,\mathbf{q}_1 + r_{22}\,\mathbf{q}_2) = \mathbf{q}_1 \cdot (A\,\mathbf{q}_2) = (A\,\mathbf{q}_1) \cdot \mathbf{q}_2 = r_{11}\,\mathbf{q}_1 \cdot \mathbf{q}_2 = 0.$$

Therefore $A\,\mathbf{q}_2 = r_{22}\,\mathbf{q}_2$ and hence $\mathbf{q}_2$ is an eigenvector with eigenvalue r_{22}. For the inductive step, in general we have

$$A\,\mathbf{q}_k = r_{1k}\mathbf{q}_1 + r_{2k}\mathbf{q}_2 + \cdots + r_{k-1,k}\mathbf{q}_{k-1} + r_{kk}\mathbf{q}_k.$$

Assuming that $\mathbf{q}_1, \ldots, \mathbf{q}_{k-1}$ are eigenvectors with eigenvalues $r_{11}, \ldots, r_{k-1,k-1}$, we take the dot product with $\mathbf{q}_j$ with $j < k$ to obtain $r_{jk} = \mathbf{q}_j \cdot A\,\mathbf{q}_k = (A\,\mathbf{q}_j) \cdot \mathbf{q}_k = r_{jj}\,\mathbf{q}_j \cdot \mathbf{q}_k = 0$. Therefore $A\,\mathbf{q}_k = r_{kk}\mathbf{q}_k$, which completes the proof. ■

7.1. Find the singular values of the following matrices and then write out their singular value decomposition: (a)♡ $(\,2, -1, 3\,)$, (b)♡ $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, (c) $\begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$, (d)♢ $\begin{pmatrix} 1 & -2 \\ -3 & 6 \end{pmatrix}$, (e)♡ $\begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix}$, (f)♢ $\begin{pmatrix} 0 & 1 \\ 1 & -1 \\ -1 & 0 \\ 1 & -1 \end{pmatrix}$, (g)♢ $\begin{pmatrix} 2 & 1 & 0 & -1 \\ 0 & -1 & 1 & 1 \end{pmatrix}$, (h) $\begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}$.

Solution: (a) $(\,1\,)(\sqrt{14})\left(\frac{2}{\sqrt{14}} \quad -\frac{1}{\sqrt{14}} \quad \frac{3}{\sqrt{14}} \right)$, (b) $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$,
(d) $\begin{pmatrix} -\frac{1}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} \end{pmatrix}(5\sqrt{2})\left(-\frac{1}{\sqrt{5}} \quad \frac{2}{\sqrt{5}} \right)$, (e) $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$.

7.4. ♡ *True or false*: If A is a symmetric matrix, then its singular values are the same as its eigenvalues.

Solution: False. The singular values are the absolute values of the nonzero eigenvalues.

7.6. ♡ Suppose Q is an orthogonal $n \times n$ matrix. What are its singular values?

Solution: $\sigma_1 = \cdots = \sigma_n = 1$.

7.9. Use the power method to find the largest singular value of the following matrices:

$$(a)\heartsuit \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}, \ (b)\diamondsuit \begin{pmatrix} 2 & 1 & -1 \\ -2 & 3 & 1 \end{pmatrix}, \ (c)\heartsuit \begin{pmatrix} 2 & 2 & 1 & -1 \\ 1 & -2 & 0 & 1 \end{pmatrix}, \ (d) \begin{pmatrix} 3 & 1 & -1 \\ 1 & -2 & 2 \\ 2 & -1 & 1 \end{pmatrix}.$$

Solution: To find the dominant singular value of a matrix A, we apply the power method to $H = A^T A$ and take the square root of its dominant eigenvalue to find the dominant singular value $\sigma_1 = \sqrt{\lambda_1}$ of A.

(a) $H = \begin{pmatrix} 2 & -1 \\ -1 & 13 \end{pmatrix}$; after 11 iterations, $\lambda_1 = 13.0902$ and $\sigma_1 = 3.6180$;

(c) $H = \begin{pmatrix} 5 & 2 & 2 & -1 \\ 2 & 8 & 2 & -4 \\ 2 & 2 & 1 & -1 \\ -1 & -4 & -1 & 2 \end{pmatrix}$; after 16 iterations, $\lambda_1 = 11.6055$ and $\sigma_1 = 3.4067$.

7.10. Compute the Euclidean matrix norm of the following matrices.

$$(a)\heartsuit \begin{pmatrix} \frac12 & \frac14 \\ \frac13 & \frac16 \end{pmatrix}, \quad (b)\diamondsuit \begin{pmatrix} \frac53 & \frac43 \\ -\frac76 & -\frac56 \end{pmatrix}, \quad (c)\heartsuit \begin{pmatrix} \frac27 & -\frac27 \\ -\frac27 & \frac67 \end{pmatrix}, \quad (d) \begin{pmatrix} \frac14 & \frac32 \\ -\frac12 & \frac54 \end{pmatrix}.$$

Solution: (a) 0.671855, (c) 0.9755.

7.12. ♡ *True or false*: The minimum value of the quantity in (5.107) is the smallest singular value of A.

Solution: True if A is nonsingular, in view of (5.50). If A is singular, the minimum value is 0, which, by our convention, is not a singular value.

7.14. ♡ Prove that the Euclidean matrix norm is bounded by the Frobenius norm, so that $\| A \|_2 \leq \| A \|_F$. When are they equal?

Solution: Using Exercise 7.13, $\| A \|_F^2 = \sigma_1^2 + \ \cdots \ + \sigma_r^2 \geq \sigma_1^2 = \| A \|_2^2$. They are equal if and only if A has only one singular value if and only if $\operatorname{rank} A = 1$. ■

7.17. Find the condition number of the following matrices. Which would you characterize as ill-conditioned?

$$(a)\heartsuit \begin{pmatrix} 2 & -1 \\ -3 & 1 \end{pmatrix}, \ (b)\diamondsuit \begin{pmatrix} 1 & 2 \\ 1.001 & 1.9997 \end{pmatrix}, \ (c)\heartsuit \begin{pmatrix} -1 & 3 & 4 \\ 2 & 10 & 6 \\ 1 & 2 & -3 \end{pmatrix}, \ (d) \begin{pmatrix} 72 & 96 & 103 \\ 42 & 55 & 59 \\ 67 & 95 & 102 \end{pmatrix}.$$

Solution: (a) The singular values are 3.8643, .2588, and so the condition number is
$3.86433 \,/\, .25878 = 14.9330$.

(c) The singular values are 12.6557, 4.34391, .98226, and so the condition number is
$12.6557 \,/\, .98226 = 12.88418$.

Chapter 6

Basics of Optimization

1.1. Find all local and global extremizers on $\mathbb{R}$ of the following scalar functions:
(a)♡ x^3-2x+1, (b)♢ $\dfrac{x}{1+x^2}$, (c)♡ $\dfrac{x^2-3x+5}{x^2+1}$, (d)♢ $e^{x^2-2x^4}$, (e) $\sin x+\frac{1}{2}\cos 2x$.

Solution: (a) Local minimizer: $x=\sqrt{\frac{2}{3}}\approx .8165,\ F\left(\sqrt{\frac{2}{3}}\right)=1-\frac{4}{9}\sqrt{6}\approx -.0887;$

local maximizer: $x=-\sqrt{\frac{2}{3}}\approx -.8165,\ F\left(-\sqrt{\frac{2}{3}}\right)=1+\frac{4}{9}\sqrt{6}\approx 2.0887;$

(c) global minimizer: $x=3,\ F(3)=\frac{1}{2}$; global maximizer: $x=-\frac{1}{3},\ F\left(-\frac{1}{3}\right)=\frac{11}{2}$.

1.2. Minimize and maximize the following objective functions on the indicated domains:
(a)♡ $x^3-2x^2+x,\ -1\le x\le 1$; (b)♢ $x^5-2x^3+x-3,\ 0\le x\le 2$;
(c)♡ $\dfrac{x^2-x}{x^2+1},\ -3\le x\le 3$; (d) $\sin(x^2+1),\ 0\le x\le 2$.

Solution:
(a) Minimum at $x=-1$ with $F(-1)=-4$; maximum at $x=\frac{1}{3}$ with $F\left(\frac{1}{3}\right)=\frac{4}{27}$

(c) minimum at $x=\sqrt{2}-1$ with $F\left(\sqrt{2}-1\right)=\dfrac{\sqrt{2}-2}{\sqrt{2}}\approx -.20711;$

maximum at $x=-\sqrt{2}-1$ with $F\left(-\sqrt{2}-1\right)=\dfrac{\sqrt{2}+2}{\sqrt{2}}\approx 1.20711.$

2.1. For each of the following quadratic functions, determine whether there is a minimizer. If so, find the minimizer and the minimum value. (a)♡ $x^2-2xy+4y^2+x-1$,
(b) $3x^2+3xy+3y^2-2x-2y+4$, (c)♡ $x^2+5xy+3y^2+2x-y$,
(d)♡ $x^2+y^2+yz+z^2+x+y-z$, (e) $x^2+xy-y^2-yz+z^2-3$,
(f)♢ $x^2+5xz+y^2-2yz+z^2+2x-z-3$, (g) $x^2+xy+y^2+yz+z^2+zw+w^2-2x-w$.

Solution: (a) Minimizer: $x=-\frac{2}{3},\ y=-\frac{1}{6}$; minimum value: $-\frac{4}{3}$. (c) No minimizer.
(d) Minimizer: $x=-\frac{1}{2},\ y=-1,\ z=1$; minimum value: $-\frac{5}{4}$.

2.3. For each matrix H, vector $\mathbf{f}$, and scalar c, write out the quadratic function $P(\mathbf{x})$ given by (6.10). Then either find the minimizer $\mathbf{x}^\star$ and minimum value $P(\mathbf{x}^\star)$, or explain why there is none.

(a) ♡ $H = \begin{pmatrix} 4 & -12 \\ -12 & 45 \end{pmatrix}$, $\mathbf{f} = \begin{pmatrix} -\frac{1}{2} \\ 2 \end{pmatrix}$, $c = 3$; (b) ♡ $H = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix}$, $\mathbf{f} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$, $c = 0$;

(c) ♡ $H = \begin{pmatrix} 3 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & -1 & 3 \end{pmatrix}$, $\mathbf{f} = \begin{pmatrix} 0 \\ 4 \\ -4 \end{pmatrix}$, $c = 6$; (d) ♢ $H = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 1 \end{pmatrix}$,

$\mathbf{f} = \begin{pmatrix} -3 \\ -1 \\ 2 \end{pmatrix}$, $c = 1$; (e) $H = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 1 & 4 \end{pmatrix}$, $\mathbf{f} = \begin{pmatrix} -1 \\ 2 \\ -3 \\ 4 \end{pmatrix}$, $c = 0$.

Solution: (a) $P(\mathbf{x}) = 2x^2 - 12xy + \frac{45}{2}y^2 + \frac{1}{2}x - \frac{7}{2}y + 3$;
minimizer: $\mathbf{x}^\star = \left(\frac{1}{24}, \frac{1}{18}\right)^T \approx (.0417, .0556)^T$; minimum value: $P(\mathbf{x}^\star) = \frac{851}{288} \approx 2.95486$.
(b) $P(\mathbf{x}) = \frac{3}{2}x^2 + 2xy + y^2 - 4x - y$; no minimizer since H is not positive definite.
(c) $P(\mathbf{x}) = \frac{3}{2}x^2 - xy + xz + y^2 - yz + \frac{3}{2}z^2 - 4y + 4z + 6$;
minimizer: $\mathbf{x}^\star = (1, 2, -1)^T$; minimum value: $P(\mathbf{x}^\star) = 0$.

2.6. ♡ Show that the quadratic function $P(x,y) = x^2 + y$ has a positive semidefinite coefficient matrix, but no minimum.

Solution: The coefficient matrix is $H = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$ and $\mathbf{x}^T H \mathbf{x} = 2x^2 \geq 0$. However, $P(0,y) = y$ can be made as large negative as desired.

2.9. ♡ Under what conditions does a quadratic function (6.10) have a finite global maximum? Explain how to find the maximizer and maximum value.

Solution: $P(\mathbf{x})$ has a maximum if and only if $-P(\mathbf{x})$ has a minimum. Thus, we require either H is negative definite, or negative semidefinite with $\mathbf{f} \in \operatorname{img} H$. The maximizer $\mathbf{x}^\star$ is obtained by solving $H\mathbf{x}^\star = \mathbf{f}$, and the maximum value $P(\mathbf{x}^\star)$ is given as before by any of the expressions in (6.14).

2.10. Find the maximum value of the quadratic functions
(a) ♡ $-x^2 + 3xy - 5y^2 - x + 1$, (b) $-2x^2 + 6xy - 3y^2 + 4x - 3y$.

Solution: (a) maximizer: $\mathbf{x}^\star = \left(\frac{10}{11}, \frac{3}{11}\right)^T$; maximum value: $p(\mathbf{x}^\star) = \frac{16}{11}$.

2.13. Find the minimizer and minimum value of the following quadratic functions when subject to the indicated constraint. (a) ♡ $x^2 - 2xy + 6y^2$, $x + y = 1$,
(b) ♢ $x^2 + y^2 + 2yz + 4z^2$, $x + 2y - z = 3$, (c) $x^2 + xy - y^2 - yz + z^2$, $x - y - z = 1$.

Solution: (a) Minimizer: $x^\star = \frac{7}{9}$, $y^\star = \frac{2}{9}$; minimum value $= \frac{5}{9}$.

2.15. ♡ Let H be a symmetric matrix. Suppose V is a subspace spanned by one or more eigenvectors of H having positive eigenvalues. Show that the restriction of the quadratic function (6.10) to V has a unique global minimum. Write down the linear system the minimum must satisfy.

Solution: Let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be an orthonormal basis of eigenvectors of H such that $\mathbf{u}_1, \ldots, \mathbf{u}_k$ have positive eigenvalues $\lambda_1, \ldots, \lambda_k > 0$. Let $\mathbf{x} = y_1 \mathbf{v}_1 + \cdots + y_k \mathbf{v}_k \in V$. Then,

$$\mathbf{x}^T H \mathbf{x} = \frac{1}{2} \sum_{i,j=1}^{k} y_i y_j \mathbf{u}_i^T H \mathbf{u}_j = \sum_{i,j=1}^{k} y_i y_j \lambda_j \mathbf{u}_i^T \mathbf{u}_j = \sum_{i=1}^{k} \lambda_i y_i^2 > 0$$

for $\mathbf{y} = (\, y_1, y_2, \ldots, y_k \,)^T \neq \mathbf{0}$, and hence the restricted quadratic function is positive definite. Further, write $\mathbf{f} = b_1 \mathbf{u}_1 + \cdots + b_n \mathbf{v}_n$ in terms of the eigenvector basis, so when $\mathbf{x} \in V$, the restricted quadratic function

$$P(\mathbf{x}) = \tfrac{1}{2} \mathbf{x}^T H \mathbf{x} - \mathbf{x}^T \mathbf{f} + c = \frac{1}{2} \sum_{i=1}^{k} \lambda_i y_i^2 - \sum_{i=1}^{k} b_i f_i + c$$

has a unique global minimum when $y_i = b_i / \lambda_i$. ■

3.1. Find the standard gradient, where it exists, of the following functions.:
(a) ♡ $x_1 x_2^2$, (b) ♢ $\log(x_1^2 + x_2^2)$, (c) ♡ $e^{x_1 - 2x_2}$, (d) $\tan^{-1}(x_1/x_2)$.

Solution: (a) $\left(x_2^2, 2 x_1 x_2 \right)^T$; (c) $\left(e^{x_1 - 2x_2}, -2\, e^{x_1 - 2x_2} \right)^T$.

3.2. Repeat Exercise 3.1 using the inner products
(i) $\langle \mathbf{x}, \mathbf{y} \rangle = 3 x_1 y_1 + 2 x_2 y_2$; (ii) $\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 - x_1 y_2 - x_2 y_1 + 4 x_2 y_2$.

Solution: (a) (i) $\left(\frac{1}{3} x_2^2, x_1 x_2 \right)^T$, (ii) $\left(\frac{2}{3} x_1 x_2 + \frac{4}{3} x_2^2, \frac{2}{3} x_1 x_2 + \frac{1}{3} x_2^2 \right)^T$
(c) (i) $\left(\frac{1}{3} e^{x_1 - 2x_2}, - e^{x_1 - 2x_2} \right)^T$, (ii) $\left(\frac{2}{3} e^{x_1 - 2x_2}, -\frac{1}{3} e^{x_1 - 2x_2} \right)^T$.

3.3. Find the critical points of the following objective functions:
(a) ♡ $x^4 + y^4 - 4xy$, (b) ♡ $xy(1 - x - y)$, (c) ♢ $x y e^{-2x^2 - 2y^2}$, (d) $(x - y)\cos y$.

Solution: (a) $(\, 0, 0 \,)^T, (\, 1, 1 \,)^T, (\, -1, -1 \,)^T$; (b) $\left(\frac{1}{3}, \frac{1}{3} \right)^T, (0,0), (0,1), (1,0)$.

3.6. Let $y = f(x)$ and $z = g(y)$ be continuously differentiable scalar functions, and let $h(x) = g \circ f(x)$ denote their composition. *True or false*:
(a) ♡ A critical point of $f(x)$ is a critical point of $h(x)$.
(b) ♢ A local minimizer of $f(x)$ is a local minimizer of $h(x)$.
(c) ♡ A critical point of $h(x)$ is a critical point of $f(x)$.
(d) A local minimizer of $h(x)$ is a local minimizer of $f(x)$.

Solution: (a) True, since if $f'(x) = 0$, then $h'(x) = g'\big(f(x)\big)\, f'(x) = 0$.

(c) False: if $g'(y_\star) = 0$ for $y_\star = f(x_\star)$ with $f'(x_\star) \neq 0$, then $h'(x_\star) = g'\big(f(x_\star)\big)\, f'(x_\star) = 0$, so $x_\star$ is a critical point for $h(x)$ but not for $f(x)$.

4.1. ♡ Write Python code to implement gradient descent on the functions $F_1(x, y) = x^2 + 2y^2$, $F_2(x, y) = x^2 + 10y^2$ and $F_3(x, y) = \sin x \, \sin y$ and numerically investigate the rates of convergence. You will need to choose the time step α by hand in each case to get the fastest convergence rate. For which function does gradient descent converge the most quickly?

Solution:

Python Notebook: Solution to Exercise 4.1 (.ipynb)

4.4. ♡ Prove that, provided $\nabla F(\mathbf{x}_k) \neq \mathbf{0}$, the inequality (6.40) holds when $\alpha_k > 0$ is sufficiently small.

Solution: Set $g(t) = F\big(\mathbf{x}_k - t\,\nabla F(\mathbf{x}_k)\big)$. By the chain rule, $g'(0) = -\,\|\,\nabla F(\mathbf{x}_k)\,\|^2 < 0$ by our assumption. Thus, $g(t) < g(0)$ for $t > 0$ sufficiently small. Replacing t by α_k completes the proof. ■

4.5. ♡ (*a*) Show that the system $x^2 + y^2 = 1, \;\; x + y = 2$, does not have a solution.
(*b*) Use gradient descent to construct a "least squares solution" by minimizing the scalar valued function $F(x, y) = (x^2 + y^2 - 1)^2 + (x + y - 2)^2$.

Solution: (*a*) Since a solution must satisfy $4 = (x + y)^2 = x^2 + 2xy + y^2 = 1 + 2xy$ and so $xy = \frac{3}{2}$, but this is not possible since $|\,x\,|, \,|\,y\,| \leq 1$.
(*b*) The least squares solution is $x = y = 1/\sqrt[3]{2} \approx .7937$.

5.1. Solve the following linear systems by the conjugate gradient method, keeping track of the residual vectors and solution approximations as you iterate.

(*a*) ♡ $\begin{pmatrix} 3 & -1 \\ -1 & 5 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, (*b*) $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} -3 \\ 1 \end{pmatrix}$, (*c*) ♡ $\begin{pmatrix} 6 & 2 & 1 \\ 2 & 3 & -1 \\ 1 & -1 & 2 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}$,

(*d*) ♢ $\begin{pmatrix} 6 & -1 & -1 & 5 \\ -1 & 7 & 1 & -1 \\ -1 & 1 & 3 & -3 \\ 5 & -1 & -3 & 6 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 0 \\ -1 \end{pmatrix}$, (*e*) $\begin{pmatrix} 5 & 1 & 1 & 1 \\ 1 & 5 & 1 & 1 \\ 1 & 1 & 5 & 1 \\ 1 & 1 & 1 & 5 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 4 \\ 0 \\ 0 \\ 0 \end{pmatrix}$.

Solution: In each, the fibal $\mathbf{x}_k$ is the actual solution, with residual $\mathbf{r}_k = \mathbf{b} - H\mathbf{x}_k \approx \mathbf{0}$.

(*a*) $\mathbf{r}_0 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, $\mathbf{x}_1 = \begin{pmatrix} .76923 \\ .38462 \end{pmatrix}$, $\mathbf{r}_1 = \begin{pmatrix} .07692 \\ -.15385 \end{pmatrix}$, $\mathbf{x}_2 = \begin{pmatrix} .78571 \\ .35714 \end{pmatrix}$; (*c*) $\mathbf{r}_0 = \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}$,

$$\mathbf{x}_1 = \begin{pmatrix} .5 \\ 0 \\ -1 \end{pmatrix}, \quad \mathbf{r}_1 = \begin{pmatrix} -1 \\ -2 \\ -.5 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} .51814 \\ -.72539 \\ -1.94301 \end{pmatrix}, \quad \mathbf{r}_2 = \begin{pmatrix} 1.28497 \\ -.80311 \\ .64249 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} 1. \\ -1.4 \\ -2.2 \end{pmatrix}.$$

5.5. ♡ Use the conjugate gradient method to solve the system $A\mathbf{u} = \mathbf{e}_5$ with coefficient matrix

$$A = \begin{pmatrix} 4 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 4 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 4 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 4 & -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 4 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 4 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1 & 4 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 4 \end{pmatrix}.$$

How many iterations do you need to obtain the solution that is accurate to 2 decimal places? *Remark*: This matrix arises in the numerical discretization of the two-dimensional Laplace partial differential equation, of great importance in many applications, [180].

Solution: Remarkably, after only two iterations, the method finds the exact solution: $\mathbf{x}_3 = \mathbf{x}^\star = (\,.0625, .125, .0625, .125, .375, .125, .0625, .125, .0625\,)^T$.

6.1. When possible, use Theorem 6.30 to determine the status of the critical points you found in Exercises 3.3 and 3.4.

Solution: Exercise 3.3: (a) $(\,1,1\,)^T, (\,-1,-1\,)^T$ are local minimizers; $(\,0,0\,)^T$ is a saddle point; (c) $(\,\frac{1}{2},-\frac{1}{2}\,)^T, (\,-\frac{1}{2},\frac{1}{2}\,)^T$ are local minimizers; $(\,\frac{1}{2},\frac{1}{2}\,)^T, (\,-\frac{1}{2},-\frac{1}{2}\,)^T$ are local maximizers; $(\,0,0\,)^T$ is a saddle point.

6.2. Let $f(x) \in \mathrm{C}^4$ be a scalar function. (a) ♡ Suppose that $f'(x^\star) = f''(x^\star) = 0$, but $f'''(x^\star) \neq 0$. Prove that $x^\star$ cannot be a local minimizer or maximizer of $f(x)$.
(b) ♢ Suppose that $f'(x^\star) = f''(x^\star) = f'''(x^\star) = 0$, while $f''''(x^\star) > 0$. Is $x^\star$ necessarily a local (*i*) maximizer, (*ii*) minimizer, (*iii*) neither, or (*iv*) cannot tell with this information alone?

Solution: (a) The third order Taylor formula says $f(x) = f(x^\star) + \frac{1}{6} f'''(x^\star)\, h^3 + R(h)$, where $h = x - x^\star$, with $R(h)/h^3 \to 0$ as $h \to 0$. Thus, if $x - x^\star$ is sufficiently small, then $f(x) - f(x^\star)$ is positive on one side of $x^\star$ and negative on the other side, and hence $t^\star$ cannot be a local minimizer or maximizer. ■

6.4. ♡ Give an example of a quadratic function $Q(x,y)$ of two variables that has no critical points. If your answer is an affine function, try harder. What can you say about the graph of $Q(x,y)$?

Solution: Let $\mathbf{x} = (\,x,y\,)^T$. Any function of the form $Q(\mathbf{x}) = \mathbf{x}^T H \mathbf{u} - 2\mathbf{x}^T\mathbf{f}$ with $\det H = 0$ and $\mathbf{f} \notin \operatorname{img} H$. For instance, $Q(x,y) = x^2 - y$. The cross-sections of the graph parallel to the x axis are parabolas with progressively lower minima as y increases.

7.2. ♡ Show that $-\log x$ is strictly convex when $x > 0$. Use this to prove $\log x \leq x - 1$ for all $x > 0$, with equality if and only if $x = 1$.

Solution: Its second derivative is $x^{-2} > 0$, proving strict convexity. Moreover, it lies strictly above its tangent line at $x = 1$, which is $1 - x \leq -\log x$, which proves the inequality with equality only holding when $x = 1$. ■

7.4. Determine whether the following functions are (i) convex; (ii) strictly convex:
(a) ♡ x, (b) ♡ x^2, (c) ♡ x^3, (d) $|x|$, (e) ♢ $|x|^3$, (f) $1/(1+x^2)$.

Solution: (a) Convex; (b) strictly convex; (c) not convex.

7.6. ♡ Let $\|\cdot\|$ be any norm on $\mathbb{R}^n$. Show that the norm function $F(\mathbf{x}) = \|\mathbf{x}\|$ is convex.

Solution: Let $0 \le t \le 1$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. By the triangle inequality,

$$F\big((1-t)\mathbf{x} + t\mathbf{y}\big) = \|(1-t)\mathbf{x} + t\mathbf{y}\| \le \|(1-t)\mathbf{x}\| + \|t\mathbf{y}\| = (1-t)F(\mathbf{x}) + tF(\mathbf{y}).$$

Therefore F is convex. ■

7.8. (a) ♡ Prove that if $F\colon \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, and both F and $-F$ are convex, then F is an affine function. (b) ♢ Is the result also valid for general functions F? If so, prove it. If not, find an explicit counterexample.

Solution: (a) If both F and $-F$ are convex, then the convexity inequality (6.91) holds for both F and $-F$, which implies that it is an equality: $F(\mathbf{y}) = F(\mathbf{x}) + \langle \nabla F(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Setting $\mathbf{x} = \mathbf{0}$ and replacing $\mathbf{y}$ by $\mathbf{x}$ yields $F(\mathbf{x}) = F(\mathbf{0}) + \langle \nabla F(\mathbf{0}), \mathbf{x} \rangle = \langle \mathbf{a}, \mathbf{x} \rangle + b$, which is an affine function with $\mathbf{a} = \nabla F(\mathbf{0})$ and $b = F(\mathbf{0})$. ■

7.10. ♡ Prove Lemma 6.37.

Solution: Let us write $H(\mathbf{x}) = aF(\mathbf{x}) + bG(\mathbf{x})$. Let $0 \le t \le 1$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then, since F and G are convex, and $a, b \ge 0$, we have

$$\begin{aligned} H((1-t)\mathbf{x} + t\mathbf{y}) &= aF((1-t)\mathbf{x} + t\mathbf{y}) + bG((1-t)\mathbf{x} + t\mathbf{y}) \\ &\le a((1-t)F(\mathbf{x}) + tF(\mathbf{y})) + b((1-t)G(\mathbf{x}) + tG(\mathbf{y})) \\ &= (1-t)(aF(\mathbf{x}) + bG(\mathbf{x})) + t(aF(\mathbf{y}) + bG(\mathbf{y})) = (1-t)H(\mathbf{x}) + tH(\mathbf{y}), \end{aligned}$$

which completes the proof. ■

7.13. ♡ Let $F\colon [0, \infty) \to \mathbb{R}$ be a convex function satisfying $F(0) = 0$. Show that F is *superadditive*, which means that $F(x) + F(y) \le F(x+y)$ for all $x, y \ge 0$.
Hint: Use (6.88) to show that $F(x) \le \dfrac{x}{x+y} F(x+y)$ and $F(y) \le \dfrac{y}{x+y} F(x+y)$.

Solution: If either x or y is zero, the inequality is trivial, so we may assume $x + y > 0$. Since $F(0) = 0$ and F is convex we have

$$F(x) = F\left(\frac{x}{x+y}(x+y) + \frac{y}{x+y}\,0\right) \le \frac{x}{x+y} F(x+y),$$

where we used definition of convexity with $t = x/(x+y)$. The inequality

$$F(y) \le \frac{y}{x+y} F(x+y)$$

follows by swapping x and y. Adding these two inequalities produces

$$F(x) + F(y) \le \frac{x}{x+y} F(x+y) + \frac{y}{x+y} F(x+y) = \frac{x+y}{x+y} F(x+y) = F(x+y).$$ ■

7.17. ♡ Prove Hölder's inequality (6.99) when $p = \infty$ and $q = 1$.

Solution: $\mathbf{x} \cdot \mathbf{y} = \displaystyle\sum_{i=1}^{n} x_i\, y_i \le \sum_{i=1}^{n} |x_i|\,|y_i| \le \left(\sum_{i=1}^{n} |x_i| \right) \max_{1 \le j \le n} |y_j| = \|\mathbf{x}\|_1\, \|\mathbf{y}\|_\infty.$ ■

7.20. ♡ Show that $F\colon \mathbb{R}^n \to \mathbb{R}$ is μ-strongly convex if and only if

$$F\big((1-t)\mathbf{x} + t\mathbf{y} \big) + \tfrac{1}{2}\mu t(1-t)\,\|\mathbf{x} - \mathbf{y}\|^2 \le (1-t)\,F(\mathbf{x}) + t\,F(\mathbf{y}) \qquad (6.112)$$

holds for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $0 \le t \le 1$.

Solution: By definition, F is μ-strongly convex if and only if $G(\mathbf{x}) = F(\mathbf{x}) - \frac{1}{2}\mu\,\|\mathbf{x}\|^2$ is convex. Then, for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $0 \le t \le 1$, the convexity inequality (6.87) for G says

$$\begin{aligned} F\big((1-t)\mathbf{x} + t\mathbf{y} \big) - \tfrac{1}{2}\mu\|(1-t)\mathbf{x} + t\mathbf{y}\|^2 &= G\big((1-t)\mathbf{x} + t\mathbf{y} \big) \le (1-t)\,G(\mathbf{x}) + t\,G(\mathbf{y}) \\ &= (1-t)\,F(\mathbf{x}) + t\,F(\mathbf{y}) - \tfrac{1}{2}\mu(1-t)\,\|\mathbf{x}\|^2 - \tfrac{1}{2}\mu t\,\|\mathbf{y}\|^2. \end{aligned}$$

Replacing $\|(1-t)\mathbf{x} + t\mathbf{y}\|^2 = (1-t)^2\,\|\mathbf{x}\|^2 + 2t(1-t)\,\mathbf{x}\cdot\mathbf{y} + t^2\,\|\mathbf{y}\|^2$, and then rearranging terms yields

$$\begin{aligned} F\big((1-t)\mathbf{x} + t\mathbf{y} \big) &\le (1-t)\,F(\mathbf{x}) + t\,F(\mathbf{y}) - \tfrac{1}{2}\mu t(1-t)\,\big[\,\|\mathbf{x}\|^2 - 2\,\mathbf{x}\cdot\mathbf{y} + \|\mathbf{y}\|^2\,\big] \\ &= (1-t)\,F(\mathbf{x}) + t\,F(\mathbf{y}) - \tfrac{1}{2}\mu t(1-t)\,\|\mathbf{x} - \mathbf{y}\|^2, \end{aligned}$$

which proves (6.112). Conversely, if F satisfies (6.112), then

$$\begin{aligned} G\big((1-t)\mathbf{x} + t\mathbf{y} \big) &= F\big((1-t)\mathbf{x} + t\mathbf{y} \big) - \tfrac{1}{2}\mu\|(1-t)\mathbf{x} + t\mathbf{y}\|^2 \\ &= F\big((1-t)\mathbf{x} + t\mathbf{y} \big) - \tfrac{1}{2}\mu(1-t)^2\,\|\mathbf{x}\|^2 - t(1-t)\,\mathbf{x}\cdot\mathbf{y} - \tfrac{1}{2}\mu t^2\,\|\mathbf{y}\|^2 \\ &= F\big((1-t)\mathbf{x} + t\mathbf{y} \big) + \tfrac{1}{2}\mu t(1-t)\,\|\mathbf{x} - \mathbf{y}\|^2 - \tfrac{1}{2}\mu(1-t)\,\|\mathbf{x}\|^2 - \tfrac{1}{2}\mu t\,\|\mathbf{y}\|^2 \\ &\le (1-t)\,F(\mathbf{x}) + t\,F(\mathbf{y}) - \tfrac{1}{2}\mu(1-t)\,\|\mathbf{x}\|^2 - \tfrac{1}{2}\mu t\,\|\mathbf{y}\|^2 = (1-t)\,G(\mathbf{x}) + t\,G(\mathbf{y}), \end{aligned}$$

and hence G is convex. ■

8.1. Determine whether or not the following scalar functions are Lipschitz continuous on $\mathbb{R}$. If so, find their Lipschitz constant.
(*a*) ♡ $|\,x\,| + |\,x - 1\,|$, (*b*) $x^{2/3}$, (*c*) ♡ $\operatorname{sign} x$, (*d*) ♡ e^x, (*e*) ♢ e^{-x^2}, (*f*) $\tanh x$.

Solution: (*a*) Lipschitz constant $= 2$; (*c*) not continuous; (*d*) not Lipschitz

8.2. Do your answers to Exercise 8.1 change if the domain is restricted to $[-1, 1]$?

Solution: (*a*) Lipschitz constant $= 2$; (*c*) not continuous; (*d*) Lipschitz constant $= e$

8.5. Determine whether or not the following functions on $\mathbb{R}^2$ are Lipschitz continuous.
(*a*) ♡ $|\,x - y\,|$, (*b*) $\max\{\,|\,x\,|, |\,y\,|\,\}$, (*c*) ♡ $x^2 - y^2$, (*d*) ♢ $\exp(-x^2 - y^2)$.

Solution: (*a*) Lipschitz continuous; (*c*) not Lipschitz continuous

8.8. ♡ Suppose the scalar-valued functions $F_1, F_2 \colon \mathbb{R}^2 \to \mathbb{R}$ are both Lipschitz continuous. (*a*) Prove that the vector-valued function $F(x,y) = (\, F_1(x,y), F_2(x,y)\,)^T$ is Lipschitz continuous. (*b*) Is the converse to part (*a*) valid?

Solution: (*a*) By the equivalence of norms, we can use the 1 norm on $\mathbb{R}^2$ to prove this. Let λ_1, λ_2 be their respective Lipschitz constants. Then, given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$,

$$\begin{aligned} \|\, F(\mathbf{x}) - F(\mathbf{y})\,\|_1 &= |\, F_1(\mathbf{x}) - F_1(\mathbf{y})\,| + |\, F_2(\mathbf{x}) - F_2(\mathbf{y})\,| \\ &\le \lambda_1 \|\,\mathbf{x}-\mathbf{y}\,\| + \lambda_2 \|\,\mathbf{x}-\mathbf{y}\,\| = \lambda \|\,\mathbf{x}-\mathbf{y}\,\|, \quad \text{where} \quad \lambda = \lambda_1 + \lambda_2. \end{aligned}$$ ■

(*b*) Yes, since for $i = 1,2$, we have $|\, F_i(\mathbf{x}) - F_i(\mathbf{y})\,| \le \|\, F(\mathbf{x}) - F(\mathbf{y})\,\|_1 \le \lambda \|\,\mathbf{x}-\mathbf{y}\,\|$.

8.10. ♡ Prove that the property of Lipschitz continuity does not depend on the underlying norm on $\mathbb{R}^n$.

Solution: Let $\|\cdot\|_a$ and $\|\cdot\|_b$ be any two norms on $\mathbb{R}^n$. If F is Lipschitz continuous on $\Omega \subset \mathbb{R}^n$ under the second norm, then

$$|\, F(\mathbf{x}) - F(\mathbf{y})\,| \le \lambda_b \|\,\mathbf{x}-\mathbf{y}\,\|_b \le \lambda_b R^\star \|\,\mathbf{x}-\mathbf{y}\,\|_a \qquad \text{for all} \qquad \mathbf{x},\mathbf{y} \in \Omega,$$

where $R^\star > 0$ is the constant appearing in the equivalence of norms inequality (2.71). Thus, F is Lipschitz continuous for the first norm, with Lipschitz constant $\lambda_a = R^\star \lambda_b$. ■

8.12. *True or false*: (*a*) ♡ A convex scalar function is Lipschitz continuous.
(*b*) ♢ A strictly convex scalar function is Lipschitz continuous.
(*c*) A Lipschitz continuous scalar function is convex.

Solution:
(*a*) False. The function $f(x) = x^2$ is convex, but not Lipschitz continuous on all of $\mathbb{R}$.

8.14. ♡ Give an example to show that (6.117) does not hold in general on connected domains. *Hint*: Take the domain to be a disk in $\mathbb{R}^2$ centered at the origin with the negative x axis removed.

Solution: As suggested, let $\Omega = \{\, x^2 + y^2 \le 1\,\} \setminus \{\, (x,0) \mid -1 \le x \le 0\,\}$. Let $F(x,y) = \theta$ be the polar angle, so $-\pi < \theta < \pi$ on Ω. The polar angle function is continuously differentiable on Ω, and even has bounded gradient if one omits a small disk centered at the origin. Let $\varepsilon > 0$ be small. Setting $\mathbf{x} = \left(-\frac{1}{2}, \varepsilon\right)^T$, $\mathbf{y} = \left(-\frac{1}{2}, -\varepsilon\right)^T$, we have $F(\mathbf{x}) - F(\mathbf{y}) \approx 2\pi$ while $\|\,\mathbf{x}-\mathbf{y}\,\| = 2\varepsilon$. Thus there is no constant $\beta > 0$ such that (6.117) holds for all $\mathbf{x},\mathbf{y} \in \Omega$.

9.1. ♡ For F_1 and F_2 from Exercise 4.1, compute the Lipschitz constant of the gradient and determine the rate of convergence of gradient descent, according to Theorem 6.66, with the optimal choice of time step. Use the Euclidean norm and dot product. Do the theoretical convergence rates match up with the experimental rates determined in Exercise 4.1?

Solution: All gradients, Hessians, and norms are Euclidean in this solution. We compute

$$\nabla F_1(x,y) = \begin{pmatrix} 2x \\ 4y \end{pmatrix}, \ \nabla F_2(x,y) = \begin{pmatrix} 2x \\ 20y \end{pmatrix}, \ \nabla^2 F_1(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}, \ \nabla^2 F_2(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 20 \end{pmatrix}.$$

By Remark 6.60, the Lipschitz constant of the gradient for each function is the largest eigenvalue of the Hessian matrix, which is 4 and 20 for F_1 and F_2, respectively. Since the optimal choice of the time step in Theorem 6.66 is $\alpha = \mathrm{Lip}(\nabla F)^{-1}$, the optimal convergence rate is $\mathrm{O}(\mathrm{Lip}(\nabla F)/k)$, which is significantly faster for F_1 compared to F_2 (5 times faster). This is the same phenomenon we observed numerically in Exercise 4.1.

9.2. ♡ Repeat Exercise 9.1, except this time compare the linear convergence rates provided by Theorem 6.68. Use the Euclidean norm and dot product.

Solution: For simplicity, assume that all gradients, Hessians, and norms are Euclidean. By (6.104) and the Hessians computed in Exercise 9.1, each F_i is μ-strongly convex with $\mu = 2$. By Remark 6.70 the convergence rate is $1 - \tau$ where τ is given by (6.146). Thus, $\tau_1 = \frac{2}{4} = \frac{1}{2}$ and $\tau_2 = \frac{2}{20} = \frac{1}{10}$, yielding again a much faster rate for F_1 as compared to F_2.

9.3. ♡ Find a preconditioning matrix C so that preconditioned gradient descent on F_2 from Exercise 4.1 is equivalent to ordinary gradient descent on F_1 (from the same exercise), and thus admits the same convergence rate.

Solution: The preconditioned gradient is $\nabla_C F = C^{-1}\nabla F$, so in order to make the preconditioned C-gradient of F_2 the same as the Euclidean gradient of F_1, we need to set $C^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{5} \end{pmatrix}$, or $C = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}$.

10.2. Use Newton's Method to find all points of intersection of the following pairs of plane curves: (*a*) ♡ $x^2 + y^2 = 1,\ xy = \frac{1}{2}$, (*b*) ♡ $x^3 + y^3 = 1,\ x^2 - y^2 = 1$,
(*c*) ♢ $x^2 + \frac{1}{3}y^2 = 1,\ x^2 + \frac{1}{4}x + 2y^2 - \frac{1}{4}y = 5$, (*d*) $y = x^2 - 3x - 5,\ x = -2y^2 + 6y$.
Hint: Sketching the curves will help you decide where to start the iterations.

Solution:

(a) $\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \approx \begin{pmatrix} 0.70710 \\ 0.70710 \end{pmatrix}$, $\begin{pmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \approx \begin{pmatrix} -0.70710 \\ -0.70710 \end{pmatrix}$. (b) $\begin{pmatrix} 0.92416 \\ 0.59505 \end{pmatrix}$, $\begin{pmatrix} 1.42503 \\ -1.23722 \end{pmatrix}$.

10.5. ♡ Prove that (6.155) is equivalent to (6.156).

Solution: Assuming $\nabla_2^2 F(\mathbf{x}_k)$ is positive definite, Theorem 6.7 implies that the minimum of the quadratic function in (6.155) is achieved when $\nabla_2^2 F(\mathbf{x}_k)\,(\mathbf{x}_{k+1} - \mathbf{x}_k) = -\nabla_2 F(\mathbf{x}_k)$. Inverting $\nabla_2^2 F(\mathbf{x}_k)$ produces the Newton iteration (6.156). ■

10.8. ♡ Given a descent direction $\mathbf{v}$ for an optimization method — for gradient descent $\mathbf{v} = -\nabla F(\mathbf{x})$, while for Newton's method $\mathbf{v} = -\nabla_2^2 F(\mathbf{x})^{-1}\nabla_2 F(\mathbf{x})$ — a *backtracking line search* aims to choose the best time step α to minimize the function F along the descent direction, that is, to minimize $F(\mathbf{x} + \alpha\mathbf{v})$ over α. The backtracking line search has two parameters $0 < \gamma \le \frac{1}{2}$ and $0 < \beta < 1$, and chooses $\alpha = \beta^k$, where $k \ge 0$ is the smallest nonnegative integer such that

$$F(\mathbf{x} + \beta^k\mathbf{v}) \le F(\mathbf{x}) + \gamma\beta^k \langle \nabla F(\mathbf{x}), \mathbf{v} \rangle. \tag{6.166}$$

In practice, one starts with $k = 0$, and then iteratively increases $k = 1, 2, \ldots$ until the inequality (6.166) holds.

(*a*) Assume ∇F is Lipschitz continuous, and the descent direction is $\mathbf{v} = -\nabla F(\mathbf{x})$. Show that there exists an integer $k \ge 0$ such that (6.166) holds. That is, the backtracking line search will eventually terminate. *Hint*: Use Lemma 6.64.

(*b*) Implement the backtracking line search in Python when $F(\mathbf{x}) = x_1^3 + 10x_2^2$. Try gradient descent, where $\mathbf{v} = -\nabla F(\mathbf{x}_k)$, and Newton's method where $\mathbf{v} = -\nabla_2^2 F(\mathbf{x}_k)^{-1}\nabla_2 F(\mathbf{x}_k)$. In both cases, after conducting the backtracking line search, the update is $\mathbf{x}_{k+1} = \mathbf{x}_k + \beta^k\mathbf{v}$. Starting from $\mathbf{x}_0 = (1, 1)^T$, you should observe faster convergence with the backtracking line search with good choices of parameters: $\gamma = 0.5$ and $\beta = 0.9$ are reasonable.

Solution: (a) By Lemma 6.64,

$$F(\mathbf{x} + \beta^k \mathbf{v}) = F(\mathbf{x} - \beta^k \nabla F(\mathbf{x})) \le F(\mathbf{x}) - \tfrac{1}{2}\,\beta^k \,\|\, \nabla F(\mathbf{x})\,\|^2,$$

provided k is large enough so that $\beta^k \le 1/\operatorname{Lip}(\nabla F)$. In contrast, the right hand side of (6.166) is

$$F(\mathbf{x}) + \gamma\,\beta^k \,\langle\, \nabla F(\mathbf{x}), \mathbf{v}\,\rangle = F(\mathbf{x}) - \gamma\,\beta^k \,\|\, \nabla F(\mathbf{x})\,\|^2.$$

The backtracking line search condition is met provided $\gamma\,\beta^k \le \frac{1}{2}\,\beta^k$, or $\gamma \le \frac{1}{2}$, which is exactly the condition on γ. ■

(*b*)

Python Notebook: Backtracking Line Search (.ipynb)

Chapter 7

Introduction to Machine Learning and Data

1.1. Find the mean, the variance, and the standard deviation of the following data sets. You can set $\nu = 1$ when computing the latter.
(a)♡ $1.1, 1.3, 1.5, 1.55, 1.6, 1.9, 2, 2.1$; (b) $2., .9, .7, 1.5, 2.6, .3, .8, 1.4$; (c)♡ $-2.9, -.5, .1, -1.5, -3.6, 1.3, .4, -.7$; (d)♢ $1.1, .2, .1, .6, 1.3, -.4, -.1, .4$; (e) $.9, -.4, -.8, .2, 1., -1.6, -1.2, -.7$.

Solution: (a) mean $= 1.63125$; variance $= .84469$; standard deviation $= 0.91907$.
(c) mean $= -0.925$; variance $= 19.375$; standard deviation $= 4.4017$.

1.2. Show that the centering matrix J is (a)♡ positive semi-definite, (b)♡ idempotent, so $J^2 = J$, (c)♢ has one-dimensional kernel spanned by $\mathbf{1}$, and hence is not positive definite, and (d) has rank $m-1$.

Solution: (a) Using the definition (7.5) to write $J\mathbf{x} = \mathbf{x} - \dfrac{1}{m}(\mathbf{1}^T\mathbf{x})\,\mathbf{1} = \mathbf{x} - \overline{x}\,\mathbf{1}$, and hence positive semidefiniteness follows from

$$\mathbf{x}^T J\,\mathbf{x} = \mathbf{x}^T\mathbf{x} - \frac{1}{m}(\mathbf{1}\cdot\mathbf{x})^2 \geq \|\,\mathbf{x}\,\|^2 - \frac{1}{m}\|\,\mathbf{1}\,\|^2\,\|\,\mathbf{x}\,\|^2 = \|\,\mathbf{x}\,\|^2 - \|\,\mathbf{x}\,\|^2 = 0,$$

where we used Cauchy–Schwarz (2.27) for the inequality, and then $\mathbf{1}^T\mathbf{1} = \|\,\mathbf{1}\,\|^2 = m$. ■

(b) $J^2 = \left(\mathrm{I} - \dfrac{1}{m}\,\mathbf{1}\,\mathbf{1}^T\right)^2 = \mathrm{I} - \dfrac{2}{m}\,\mathbf{1}\,\mathbf{1}^T + \dfrac{1}{m^2}\,\mathbf{1}\,\mathbf{1}^T\mathbf{1}\,\mathbf{1}^T = \mathrm{I} - \dfrac{1}{m}\,\mathbf{1}\,\mathbf{1}^T = J.$ ■

1.6.♡ Suppose we have a collection of data points $\mathbf{x}_1, \ldots, \mathbf{x}_m$ lying along a line spanned by the unit vector $\mathbf{u}$, that is each $\mathbf{x}_i = s_i\mathbf{u}$ for some $s_i \in \mathbb{R}$. Show that the covariance matrix of this data is $S_X = \sigma_s^2\,\mathbf{u}\,\mathbf{u}^T$, where σ_s^2 is the variance of the weights $\mathbf{s} = (s_1, \ldots, s_m)$.

Solution: Let $\bar{s} = \dfrac{1}{m}\displaystyle\sum_{i=1}^{m} s_i$ and note that $\overline{\mathbf{x}} = \bar{s}\,\mathbf{u}$. Thus, by (7.22) we have

$$S_X = \nu\sum_{i=1}^{m}(s_i - \bar{s})^2\,\mathbf{u}\,\mathbf{u}^T = \sigma_s^2\,\mathbf{u}\,\mathbf{u}^T, \qquad \text{since} \qquad \mathbf{x}_i - \overline{\mathbf{x}} = (s_i - \bar{s})\,\mathbf{u}.$$ ■

2.2. ♡ This exercise compares the 1 norm and 2 norm in regression.

(*a*) Show that the solution of the optimization problem $\min\{\,\|\,\mathbf{w}\,\|_2 \mid \mathbf{1}\cdot\mathbf{w}=1\,\}$ is given by $\mathbf{w}=\mathbf{1}/n$. This shows that under a constraint on the total *mass* of the weights, i.e., $\mathbf{w}\cdot\mathbf{1}=1$, the 2 norm prefers to assign weights equally across all features. *Hint*: Write $\mathbf{z}=\mathbf{w}-\mathbf{1}/n$ and convert it into the equivalent optimization problem $\min\{\,\|\,\mathbf{z}+\mathbf{1}/n\,\|_2 \mid \mathbf{1}\cdot\mathbf{z}=0\,\}$ whose optimal solution is $\mathbf{z}=\mathbf{0}$.

(*b*) Show that the solution of the optimization problem $\min\{\,\|\,\mathbf{w}\,\|_1 \mid \mathbf{1}\cdot\mathbf{w}=1\,\}$ is any vector $\mathbf{w}$ satisfying the constraint $\mathbf{1}\cdot\mathbf{w}=1$ that has nonnegative entries, i.e., $w_i\geq 0$ for all i. Hence, the 1 norm does not place any preference on how the mass is distributed among features, and the sparse solution $\mathbf{w}=\mathbf{e}_1$ is as equally preferred as the nonsparse solution $\mathbf{w}=\mathbf{1}/n$. This fact allows lasso and elastic net to find sparser solutions when they exist.

Solution: (a) First note that

$$1=(\mathbf{1}\cdot\mathbf{w})^2=\sum_{i=1}^{n} w_i^2+\sum_{i<j=1}^{2} w_i\,w_j.$$

Since $0\leq (w_i-w_j)^2=w_i^2+w_j^2-2\,w_iw_j$, we have $2\,w_iw_j\leq w_i^2+w_j^2$, with equality if and only if $w_i=w_j$. Using this in the first formula yields

$$1\leq n\sum_{i=1}^{n} w_i^2=n\,\|\,\mathbf{w}\,\|^2. \qquad \text{Thus,}\quad \|\,\mathbf{w}\,\|^2\geq\frac{1}{n}=\left\|\,\frac{\mathbf{1}}{n}\,\right\|^2 \quad\text{whenever}\quad \mathbf{1}\cdot\mathbf{w}=1.$$

Moreover, equality is achieved if and only if all entries of $\mathbf{w}$ are equal, so $\mathbf{w}=\mathbf{1}/n$. ■

(*b*) We have $1=w_1+\cdots+w_n\leq|\,w_1\,|+\cdots+|\,w_n\,|=\|\,\mathbf{w}\,\|_1$, with equality if and only if all $w_i\geq 0$, and hence any such $\mathbf{w}$ minimizes the one norm. ■

2.3. ♡ Consider the general ridge regression problem $\min_{\mathbf{w}}\{\,\|\,X\mathbf{w}-\mathbf{y}\,\|^2+\lambda\,\|\,B\mathbf{w}\,\|^2\,\}$. Show that the solution is unique and is given by (7.41) when $X^TX+\lambda B^TB$ is nonsingular. Explain how this solves the ridge regression problem (7.42).

Solution: Expanding,

$$\|\,X\mathbf{w}-\mathbf{y}\,\|^2+\lambda\,\|\,B\,\mathbf{w}\,\|^2=\mathbf{w}^T(X^TX+\lambda B^TB)\,\mathbf{w}-2\,\mathbf{w}^TX^T\mathbf{y}+\|\,\mathbf{y}\,\|^2.$$

This is a quadratic function of $\mathbf{w}$, and, if $\lambda>0$, the coefficient matrix $X^TX+\lambda B^TB$ is a combination of Gram matrices, and hence positive semidefinite. Thus, provided this matrix is nonsingular and hence positive definite, Theorem 6.7 implies that it has a unique global minimizer, which is given by (7.41). ■

Solution: (a) Let $X=(\mathbf{v}_1\ \ldots\ \mathbf{v}_n)$. The assumption in the problem is that $\mathbf{v}_1=\cdots=\mathbf{v}_k$. In this case, the least squares part of the ridge regression loss can be written as

$$\|\,X\mathbf{w}-\mathbf{y}\,\|^2=\left\|\,\sum_{i=1}^{n} w_i\mathbf{v}_i-\mathbf{y}\,\right\|^2=\left\|\,(w_1+\cdots+w_k)\,\mathbf{v}_1+\sum_{i=k+1}^{n} w_i\mathbf{v}_i-\mathbf{y}\,\right\|^2,$$

while the regularizer is $\lambda\,\|\,\mathbf{w}\,\|^2=\lambda(w_1^2+\cdots+w_n^2)$, where $\lambda>0$. Let $\mathbf{w}$ be the minimizer of the ridge regression problem, and set $\widetilde{\mathbf{w}}=(w,\ldots,w,w_{i+1},\ldots,w_n)$, where $w=(w_1+\cdots+w_k)/k$. Then $\widetilde{w}_1+\cdots+\widetilde{w}_k=w_1+\cdots+w_k$, which implies that $\|\,X\mathbf{w}-\mathbf{y}\,\|^2=\|\,X\widetilde{\mathbf{w}}-\mathbf{y}\,\|^2$. Note that, by the Cauchy-Schwarz inequality,

$$k^2w^2=(w_1+\cdots+w_k)^2=(\mathbf{1}\cdot\mathbf{w})^2\leq\|\,\mathbf{1}\,\|^2\,\|\,\mathbf{w}\,\|^2=k\,(w_1^2+\cdots+w_k^2).$$

It follows that

$$\widetilde{w}_1^2 + \cdots + \widetilde{w}_k^2 = k\,w^2 \le w_1^2 + \cdots + w_k^2,$$

and hence $\|\widetilde{\mathbf{w}}\|^2 \le \|\mathbf{w}\|^2$. Thus, $\widetilde{\mathbf{w}}$ is also a minimizer of ridge regression problem. Since the minimizer is unique we have $\widetilde{\mathbf{w}} = \mathbf{w}$, and so the minimizer must set $w_1 = \cdots = w_k = w$. ■

(b) We again write the least squares part of the lasso loss in the above form. Without loss of generality, we may assume that $w_1 + \cdots + w_k \ge 0$, the proof being similar in the negative case. We define $\widetilde{\mathbf{w}}$ in the same way as in part (a), and note that $\widetilde{w}_1 = \cdots = \widetilde{w}_k = w \ge 0$. Furthermore,

$$|\widetilde{w}_1| + \cdots + |\widetilde{w}_k| = w + \cdots + w = k\,w = w_1 + \cdots + w_k \le |w_1| + \cdots + |w_k|,$$

from which it follows that $\|\widetilde{\mathbf{w}}\|_1 \le \|\mathbf{w}\|_1$. Therefore $\widetilde{\mathbf{w}}$ is also a minimizer of the lasso regression problem, and its first k weights all have the same sign (they are all nonnegative). They are also all exactly the same, as in part (a), except in this case the solution of lasso is not unique, since the loss function is not strongly or strictly convex, and so we cannot conclude that $\mathbf{w} = \widetilde{\mathbf{w}}$. ■

(c) We have essentially already shown this in the previous parts. By the displayed equation in part (a), the least squares portion of the lasso regression loss is the same for $\mathbf{w}$ and $\widetilde{\mathbf{w}}$. The regularizer term is also the same, since

$$|w_1| + \cdots + |w_k| = w_1 + \cdots + w_k = \widetilde{w}_1 + \cdots + \widetilde{w}_k = |\widetilde{w}_1| + \cdots + |\widetilde{w}_k|.$$ ■

2.6. ♡ Prove that (7.53) holds.

Solution: For any $i = 1, \ldots, n$,

$$\begin{aligned}\|X\mathbf{w} - \mathbf{y}\|^2 + \lambda\,\|\mathbf{w}\|_1 &= \left\| w_i\mathbf{v}_i - \left(\mathbf{y} - \sum_{j\ne i} w_j\mathbf{v}_j\right)\right\|^2 + \lambda|w_i| + \lambda\sum_{j\ne i}|w_j| \\ &= \|\mathbf{v}_i\|^2\,w_i^2 - 2\,w_i\,b_i + \lambda\,|\,w_i\,| + c,\end{aligned}$$

where

$$b_i = \mathbf{v}_i \cdot \left(\mathbf{y} - \sum_{j\ne i} w_j\mathbf{v}_j\right) \qquad \text{and} \qquad c = \|\mathbf{y} - \sum_{j\ne i} w_j\mathbf{v}_j\|^2 + \lambda\sum_{j\ne i}|w_j|.$$

2.9. ♡ Write Python code that solves the lasso regression problem using the iterative shrinkage–thresholding algorithm (ISTA). Test your code on the diabetes data set.

Solution:

Python Notebook: Lasso via ISTA (.ipynb)

2.11. Let $x_1, \ldots, x_{k+1} \in \mathbb{R}$ be distinct real numbers, i.e., $x_i \neq x_j$ when $i \neq j$. (*a*) ♡ Prove that the corresponding $(k+1) \times (k+1)$ Vandermonde matrix (7.56) is nonsingular. *Hint*: Prove that $\ker Z = \{\mathbf{0}\}$, by using the fact that a nonzero polynomial of degree k can have at most k roots. (*b*) ♢ Given data points $y_1, \ldots, y_{k+1} \in \mathbb{R}$, an *interpolating polynomial* $p(x)$ satisfies $p(x_i) = y_i$ for all $i = 1, \ldots, k+1$. Prove that there exists a unique interpolating polynomial of degree $\leq k$ for any collection of data points. *Hint*: Write the interpolation conditions in vectorial form using the Vandermonde matrix.

Solution: (a) Suppose $\mathbf{c} = (\,c_0, c_1, \ldots, c_k\,)^T \in \ker Z$. Define the degree k polynomial $p(x) = c_0 + c_1 x + \cdots + c_k x^k$. Then the condition $Z\,\mathbf{c} = \mathbf{0}$ implies $p(x_i) = 0$ for $i = 1, \ldots, k+1$ which, by the hint, implies $p(x) \equiv 0$ and hence $c_0 = c_1 = \cdots = c_k = 0$. Thus $\ker Z = \{\mathbf{0}\}$. ■

3.1. ♡ Let $F\colon \mathbb{R}^n \to \mathbb{R}^2$, so $F(\mathbf{x}) = (\,F_1(\mathbf{x}), F_2(\mathbf{x})\,)^T$, be the output of a binary classifier, which we assume to be a probability vector. Explain why we only need to learn the scalar-valued function $F_1(\mathbf{x})$ and threshold at $F_1(\mathbf{x}) = 0.5$ to perform binary classification.

Solution: Since $F(\mathbf{x})$ is a probability vector, we have $F_1(\mathbf{x}) + F_2(\mathbf{x}) = 1$ and $F_i(\mathbf{x}) \geq 0$ for $i = 1, 2$. Given $F_1(\mathbf{x})$, we can deduce $F_2(\mathbf{x}) = 1 - F_1(\mathbf{x})$. Moreover, $F(\mathbf{x})$ is closer to $\mathbf{e}_1$ if $0 \leq F_1(\mathbf{x}) < 0.5$ and closer to $\mathbf{e}_2$ if $0.5 < F_1(\mathbf{x}) < 1$.

3.2. ♡ Given the data matrix $X = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$ and labels $\mathbf{y} = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$ find three separating hyperplanes, and find the maximal margin SVM classifier.

Solution: Any line of the form $x_1 + x_2 = c$ for $0 < c < 1$ is a separating hyperplane. The maximum margin classifier is obtained by setting $c = 0.5$, which makes the line equidistant from all points.

3.3. ♡ Use `sklearn` in Python to train a soft-margin SVM to be applied to the data matrix $X = \begin{pmatrix} 0 & 0 & 0 \\ 1 & -1 & 1 \\ 0 & 1 & 1 \\ 2 & -2 & 3 \end{pmatrix}$ and labels $\mathbf{y} = \begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \end{pmatrix}$. What are the values for $\mathbf{w}$ and b? The Python notebook for this section will be helpful, but keep in mind that, in Python, the labels must be nonnegative, so use $\mathbf{y} = (0, 1, 0, 1)^T$ and adapt your result.

Solution: A Python notebook is provided below, and the found values, which perfectly separate the data, are $\mathbf{w} = (0.6664919, -0.66675405, 0.66622976)^T$ and $b = 0.99965047$.

Python Notebook: SVM Toy Homework Problem (.ipynb)

3.8. ♡ Consider a linearly separable data set, where there exists a solution $(\mathbf{w}_0, b_0)$ of the hard-margin SVM problem (7.62). Let $(\mathbf{w}_\lambda, b_\lambda)$ be a solution of the soft-margin problem (7.64) for $\lambda > 0$. (*a*) Show that $\|\mathbf{w}_\lambda\| \le \|\mathbf{w}_0\|$. (*b*) Show that $y_i(\mathbf{x}_i \cdot \mathbf{w}_\lambda - b_\lambda) + \lambda\, m\, \|\mathbf{w}_0\|^2 \ge 1$. Therefore, in the linearly separable case, the soft-margin SVM problem with small λ provides a good approximation to the solution of the hard-margin problem.

Solution: We'll prove both parts at the same time. Since $(\mathbf{w}_\lambda, b_\lambda)$ minimizes the soft-margin SVM problem,

$$\lambda\|\mathbf{w}_\lambda\|^2 + \frac{1}{m}\sum_{i=1}^m \big[\,1 - y_i\,(\mathbf{x}_i \cdot \mathbf{w}_\lambda - b_\lambda)\,\big]_+ \le \lambda\,\|\mathbf{w}_0\|^2 + \frac{1}{m}\sum_{i=1}^m \big[\,1 - y_i\,(\mathbf{x}_i \cdot \mathbf{w}_0 - b_0)\,\big]_+.$$

Since $(\mathbf{w}_0, b_0)$ solves the hard-margin SVM problem, $y_i\,(\mathbf{x}_i \cdot \mathbf{w}_0 - b_0) \ge 1$ for all i, and so

$$\frac{1}{m}\sum_{i=1}^m [1 - y_i\,(\mathbf{x}_i \cdot \mathbf{w}_0 - b_0)]_+ = 0.$$

Therefore

$$\lambda\,\|\mathbf{w}_\lambda\|^2 + \frac{1}{m}\sum_{i=1}^m \big[\,1 - y_i\,(\mathbf{x}_i \cdot \mathbf{w}_\lambda - b_\lambda)\,\big]_+ \le \lambda\,\|\mathbf{w}_0\|^2.$$

Since all terms on the left hand side are nonnegative, we deduce that $\lambda\,\|\mathbf{w}_\lambda\|^2 \le \lambda\,\|\mathbf{w}_0\|^2$, which proves part (a), and

$$\frac{1}{m}\big[\,1 - y_i\,(\mathbf{x}_i \cdot \mathbf{w}_\lambda - b_\lambda)\,\big]_+ \le \lambda\,\|\mathbf{w}_0\|^2,$$

which upon rearranging, proves part (b). ■

4.1. ♡ Let $x_1, \ldots, x_m \in \mathbb{R}$ be a collection of one dimensional data points. Show that we can compute the nearest neighbor of every point x_i in the data set in $\mathrm{O}(m \log m)$ operations, compared to the $\mathrm{O}(m^2)$ operations it would take to compare distances between every pair of data points $|\,x_i - x_j\,|$.
Hint: Recall that the computational complexity of sorting m numbers is $\mathrm{O}(m \log m)$ [47].

Solution: We first sort the numbers in ascending order, which takes $\mathrm{O}(m \log m)$ operations, and we can henceforth assume $x_1 \le x_2 \le \cdots \le x_m$. Then the nearest neighbor of x_i is either x_{i-1} or x_{i+1}, unless $i = 1$ or $i = m$, in which case there is only one choice. Thus, we only need to check two points for each x_i, which is an additional $2m$ operations, yielding a total of $\mathrm{O}(m \log m)$ operations. ■

4.5. ♡ Let $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{R}^n$ and $\mathbf{y}_1, \ldots, \mathbf{y}_m \in \mathbb{R}^c$ denote the training data for a k-nearest neighbor classifier. Define

$$G(\mathbf{x}) = \begin{cases} 1, & \|\mathbf{x}\| \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Show that there exists a function $H\colon \mathbb{R}^n \to \mathbb{R}$, which may depend on the training data, such that the classification decision of a uniformly weighted k-nearest neighbor classifier using the norm $\|\cdot\|$ can be deduced from the function

$$F(\mathbf{x}) = \sum_{i=1}^{m} G\left(\frac{\mathbf{x} - \mathbf{x}_i}{H(\mathbf{x})}\right) \mathbf{y}_i. \tag{7.72}$$

How does the classification decision relate to $F(\mathbf{x})$? You can assume the k-th nearest neighbor of $\mathbf{x}$ is unique, so no ties have to be broken.

Solution: If we set $H(\mathbf{x})$ to be the distance from $\mathbf{x}$ to its k-th nearest neighbor, then $F(\mathbf{x})$ is the sum of the label vectors corresponding to each of these neighbors, which, up to a factor of $1/k$, is the uniformly weighted k-nearest neighbor classifier (provided no ties have to be broken). The normalization by $1/k$ does not affect the label decision, which is simply taken as the largest component of $F(\mathbf{x})$. ■

5.1. ♡ Given a data set with m points, prove that there are $2^{m-1} - 1$ possible ways to cluster the data into 2 nonempty clusters. *Remark*: The generalization of this result to k clusters is provided by the Stirling numbers of the second kind, cf. [92].

Solution: Without loss of generality, we can assume the first data point is in the first cluster; otherwise just relabel the clusters. Each of the other $m - 1$ data points can be either in cluster 1 or 2, leading to 2^{m-1} different possible clusterings. Except one of them has all the data points in the first cluster while the second cluster is empty, so we must subtract 1 from this total. ■

5.4. ♡ Consider Exercise 5.3 in dimension $n = 1$ and assume $c_1 < c_2$. Show that

$$C_1 = \left\{ x \;\middle|\; x \le \tfrac{1}{2}(c_1 + c_2) \right\}, \qquad C_2 = \left\{ x \;\middle|\; x > \tfrac{1}{2}(c_1 + c_2) \right\}.$$

Solution: In this case the inequality for C_1 reduces to

$$x\,(c_2 - c_1) = x\,w \le b = \tfrac{1}{2}\,(c_2^2 - c_1^2) = \tfrac{1}{2}\,(c_2 - c_1)\,(c_2 + c_1).$$

Dividing by $c_2 - c_1 > 0$ completes the proof. The proof for C_2 is the same calculation with $>$ replacing $\le$. ■

5.6. ♡ Complete the proof of Lemma 7.11 by showing that C_1^{t+1} is nonempty.

Solution: Assume that $C_1^{t+1} = \varnothing$, which means that $\mathbf{x}_i \cdot \mathbf{w}^{t+1} > b^{t+1}$ for all $i = 1, \ldots, m$. Then we would have that $\mathbf{c}_2^{t+1} \cdot \mathbf{w}^{t+1} > b^{t+1}$, which contradicts the fact that $\mathbf{c}_2^{t+1} \cdot \mathbf{w}^{t+1} \le b^{t+1}$, which holds by definition of $\mathbf{w}^{t+1}$ and b^{t+1}. Thus, we must have $C_2^{t+1} \ne \varnothing$. ■

5.9. ♡ (Robust k-means clustering) The exercise is focused on the robust k-means algorithm, which is guided by minimizing (7.82). We start with distinct randomized initial values for the means $\mathbf{c}_1^0, \mathbf{c}_2^0, \ldots, \mathbf{c}_k^0$ chosen from the data set, and iterate the steps below until convergence.

(i) Update the clusters as in (7.76).

(ii) Update the cluster centers

$$\mathbf{c}_j^{t+1} \in \operatorname*{argmin}_{\mathbf{c}} \sum_{\mathbf{x}\in C_j^t} \|\mathbf{x}-\mathbf{c}\|. \tag{7.85}$$

(a) Show that the robust k-means algorithm descends on the energy E_{robust}.

(b) The cluster center $\mathbf{c}_j^{t+1}$ does not admit a closed form expression and is sometimes inconvenient to work with in practice. Consider changing the Euclidean norm in (7.82) to the 1 *norm* and redefine E_{robust} as

$$E_{\text{robust}}(\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_k) = \sum_{i=1}^{m} \min_{1\le j\le k} \|\mathbf{c}_j - \mathbf{x}_i\|_1. \tag{7.86}$$

This is called *k-medians clustering*. Formulate both steps of the k-medians algorithm so that it descends on the k medians clustering energy (7.86). In particular, show that the cluster centers $\mathbf{c}_j^{t+1}$ are the coordinatewise medians of the points $\mathbf{x}\in C_j^t$, which are simple to compute.

(c) Can you think of any reasons why the Euclidean norm would be preferred over the 1 norm in the k-means energy?

(d) *Challenge*: Implement the robust k-medians algorithm in Python.

Solution: (a) The proof is very similar to Theorem 7.13. We compute

$$\begin{aligned} E_{\text{robust}}(\mathbf{c}_1^t, \ldots, \mathbf{c}_k^t) &= \sum_{j=1}^{k} \sum_{\mathbf{x}\in C_j^t} \|\mathbf{c}_j^t - \mathbf{x}\| = \sum_{j=1}^{k} \sum_{\mathbf{x}\in C_j^t} \|\mathbf{c}_j^{t+1} - \mathbf{x}\| \\ &\ge \sum_{j=1}^{k} \sum_{\mathbf{x}\in C_j^t} \min_{1\le \ell\le k} \|\mathbf{c}_\ell^{t+1} - \mathbf{x}\| = E_{\text{robust}}(\mathbf{c}_1^{t+1}, \ldots, \mathbf{c}_k^{t+1}), \end{aligned}$$

and hence the energy decreases at each step. ■

(b) The only change in the k-medians algorithm is the cluster centers step (7.85), which in this case amounts to minimizing

$$\sum_{\mathbf{x}\in C_j^t} \|\mathbf{x}-\mathbf{c}\|_1 = \sum_{\mathbf{x}\in C_j^t} \sum_{i=1}^{n} |x_i - c_i| = \sum_{i=1}^{n} \sum_{\mathbf{x}\in C_j^t} |x_i - c_i|.$$

over $\mathbf{c} = (c_1, c_2, \ldots, c_n)^T$. Hence, the coordinate c_i should be chosen to minimize

$$\sum_{\mathbf{x}\in C_j^t} |x_i - c_i|.$$

We claim that an optimal c_i is the median of the i-th coordinates of points in C_j^t. This is equivalent to showing that a solution c of

$$\min_c f(c) := \sum_{i=1}^{m} |a_i - c|$$

for any given real numbers $a_i \in \mathbb{R}$ is the median of $a_1, \ldots, a_m$. To see this, we may, without loss of generality, assume that $a_1 \le \cdots \le a_m$. We can then differentiate f at any $a_j < c < a_{j+1}$

$$f'(c) = \frac{d}{dc}\left[\sum_{i=1}^{j}(c - a_i) + \sum_{i=j+1}^{m}(a_i - c)\right] = j - (m - j) = 2j - m.$$

Thus, f is decreasing when $j \le m/2$ and increasing when $j \ge m/2$, which implies that any median is a minimizer of f. ■

(c) Unlike the Euclidean norm, the 1 norm is not invariant under rotations of Euclidean space, so you may get different clusterings upon rotating the data, which may not be desirable.

(d)

Python Notebook: k medians solution (.ipynb)

6.1. ♡ Consider the inner product feature map kernel $\mathcal{K}_\phi(\mathbf{x}, \mathbf{y}) = \langle\, \phi(\mathbf{x}), \phi(\mathbf{y})\,\rangle$. Show that there exists a feature map ψ such that $\mathcal{K}_\phi(\mathbf{x}, \mathbf{y}) = \psi(\mathbf{x}) \cdot \psi(\mathbf{y})$, so that there is no loss of generality in using the dot product in the definition of feature map kernels.

Solution: If C is the symmetric positive definite matrix that defines the inner product, then we have $\mathcal{K}_\phi(\mathbf{x}, \mathbf{y}) = \langle\, \phi(\mathbf{x}), \phi(\mathbf{y})\,\rangle = \phi(\mathbf{x})^T C\, \phi(\mathbf{y})$. Let $S = C^{1/2}$ be its square root, which is also symmetric positive definite. This allows us to write

$$\mathcal{K}_\phi(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x})^T S^2 \phi(\mathbf{y}) = \big(S\,\phi(\mathbf{x})\big)^T \big(S\,\phi(\mathbf{y})\big) = \big(S\,\phi(\mathbf{x})\big) \cdot \big(S\,\phi(\mathbf{y})\big) = \psi(\mathbf{x}) \cdot \psi(\mathbf{y}),$$

where $\psi(\mathbf{x}) = S\,\phi(\mathbf{x})$. ■

6.4. ♡ Let $\mathcal{K}$ be the sigmoid kernel function. (a) Find $\mathbf{x}$ so that $\mathcal{K}(\mathbf{x}, \mathbf{x}) < 0$ when $\kappa < 0$ or $c < 0$. (b) Show that $\mathcal{K}$ is not a Mercer kernel even when $c > 0$ and $\kappa > 0$. *Hint*: Look for an $m = 2$ counterexample.

Solution: (a) If either $\kappa < 0$ or $c < 0$, then we can find $\mathbf{x}$ such that $\kappa\,\|\,\mathbf{x}\,\|^2 + c < 0$ and hence $\mathcal{K}(\mathbf{x}, \mathbf{x}) = \tanh\big(\kappa\,\|\,\mathbf{x}\,\|^2 + c\big) < 0$. ($b$) Let $\mathbf{x}_1$ be a unit vector, $\|\,\mathbf{x}_1\,\| = 1$, and let $\mathbf{x}_2 = a\,\mathbf{x}_1$. Then

$$\begin{aligned}\det\begin{pmatrix}\mathcal{K}(\mathbf{x}_1, \mathbf{x}_1) & \mathcal{K}(\mathbf{x}_1, \mathbf{x}_2)\\ \mathcal{K}(\mathbf{x}_1, \mathbf{x}_2) & \mathcal{K}(\mathbf{x}_2, \mathbf{x}_2)\end{pmatrix} &= \det\begin{pmatrix}\tanh(\kappa + c) & \tanh(\kappa\, a + c)\\ \tanh(\kappa\, a + c) & \tanh(\kappa\, a^2 + c)\end{pmatrix}\\ &= \tanh(\kappa + c)\, \tanh(\kappa\, a^2 + c) - \tanh(\kappa\, a + c)^2.\end{aligned}$$

Since $\tanh t < 1$ for all t, and $\lim_{t\to\infty} \tanh t = 1$, in the limit as $a \to \infty$, the above determinant has limiting value $\tanh(\kappa + c) - 1 < 0$, which implies it is negative for a sufficiently large. This implies that the corresponding 2×2 sigmoid kernel matrix cannot be positive semidefinite. ■

6.7. ♡ Show that $\mathbf{c}$ defined by (7.102) is the minimal Euclidean norm solution of (7.104) when $\mathbf{y} \in \operatorname{img} K$. What happen when $\mathbf{y} \notin \operatorname{img} K$?

Solution: By the discussion leading to (7.102), any solution of (7.104) is given by

$$\mathbf{c} = (K + \lambda\, \mathrm{I}\,)^{-1}\,\mathbf{y} + \mathbf{v},$$

where $\mathbf{v} \in \ker K$. If $\mathbf{y} \in \operatorname{img} K$, then $\mathbf{y} = K\mathbf{z}$ and so

$$(K + \lambda\, \mathrm{I}\,)^{-1}\,\mathbf{y} = (K + \lambda\, \mathrm{I}\,)^{-1}\,K\mathbf{z} = K(K + \lambda\, \mathrm{I}\,)^{-1}\mathbf{z},$$

and so the first term in the definition of $\mathbf{c}$ above belongs to $\operatorname{img} K$ as well. Since $\operatorname{img} K$ and $\ker K$ are orthogonal complements, we have

$$\|\,\mathbf{c}\,\|^2 = \|\,(K + \lambda\, \mathrm{I}\,)^{-1}\,\mathbf{y}\,\|^2 + \|\,\mathbf{v}\,\|^2,$$

hence the minimal Euclidean norm solution is obtained by setting $\mathbf{v} = 0$. When $\mathbf{y} \notin \operatorname{img} K$, the first term $(K + \lambda\, \mathrm{I}\,)^{-1}\,\mathbf{y}$ has a non-zero (orthogonal) projection $\mathbf{p}$ onto $\ker K$. Setting $\mathbf{v} = -\mathbf{p}$ results in a smaller Euclidean norm, and so $\mathbf{c}$ given by (7.102) is not the minimal Euclidean norm solution in this case. ■

Chapter 8

Principal Component Analysis

1.1. ♡ Construct the 5×5 covariance matrix for the data set from Exercise 1.1 in Chapter 7, and find its principal variances, principal standard deviations, and principal directions. What do you think is the dimension of the subspace the data lies in?

Solution: Covariance matrix:

$$S_X = \begin{pmatrix} 0.844687 & -0.62375 & 2.09625 & -0.71 & -1.48875 \\ -0.62375 & 3.995 & -8.625 & 3.01 & 4.675 \\ 2.09625 & -8.625 & 19.375 & -6.74 & -10.865 \\ -0.71 & 3.01 & -6.74 & 2.36 & 3.77 \\ -1.48875 & 4.675 & -10.865 & 3.77 & 6.295 \end{pmatrix}$$

Principal variances: $1017.39,\ \ 0.8656,\ \ 0.00086,\ \ 0.000178,\ \ 0.0;$
Principal standard deviations: $31.8966,\ \ 0.93037,\ \ 0.02938,\ \ 0.01335,\ \ 0.0;$
Principal directions:

$$\begin{pmatrix} 0.08677 \\ -0.34556 \\ 0.77916 \\ -0.27110 \\ -0.43873 \end{pmatrix}, \quad \begin{pmatrix} -0.80181 \\ -0.44715 \\ 0.086880 \\ -0.056296 \\ 0.38267 \end{pmatrix}, \quad \begin{pmatrix} 0.46356 \\ -0.057289 \\ 0.31273 \\ -0.18453 \\ 0.80621 \end{pmatrix}, \quad \begin{pmatrix} -0.067789 \\ 0.13724 \\ -0.29037 \\ -0.94302 \\ -0.05448 \end{pmatrix}, \quad \begin{pmatrix} 0.36067 \\ -0.81150 \\ -0.45084 \\ 0.0 \\ -0.09017 \end{pmatrix}.$$

The data lies close to a two-dimensional subspace spanned by the first two principal directions, and is precisely on the four-dimensional subspace spanned by the first four principal directions.

1.2. ♡ Using the Euclidean norm, compute a fairly dense sample of points on the unit sphere $S = \{\,\mathbf{x} \in \mathbb{R}^3 \mid \|\mathbf{x}\| = 1\,\}$. (*a*) Set $\mu = .95$ in (8.3), and then find the principal components of your data set. Do they indicate the two-dimensional nature of the sphere? If not, why not? (*b*) Now look at the subset of your data that is within a distance $r > 0$ of the north pole, i.e., $\|\mathbf{x} - (\,0,0,1\,)^T\| \le r$, and compute its principal components. How small does r need to be to reveal the actual dimension of S? Interpret your calculations.

Solution:

Python Notebook: Solution to Exercise 1.2. (.ipynb)

1.4. ♡ Show that the first principal direction $\mathbf{q}_1$ can be characterized as the direction of the line that minimizes the sums of the squares of its distances to the data points. *Hint*: Use Theorem 5.43.

Solution: Given a line in the direction of a unit vector $\mathbf{u}$, the sum of the squares of its distances to the data points is

$$\sum_{i=1}^{m} \|\mathbf{x}_i - (\mathbf{x}_i \cdot \mathbf{u})\,\mathbf{u}\,\|^2 = \sum_{i=1}^{m} \left[\,\|\mathbf{x}_i\|^2 - (\mathbf{x}_i \cdot \mathbf{u})^2\,\right] = \sum_{i=1}^{m} \|\mathbf{x}_i\|^2 - \|\,X\,\mathbf{u}\,\|^2.$$

This is minimized by maximizing the second term,

$$\|\,X\,\mathbf{u}\,\|^2 = \mathbf{u}^T (X^T X)\,\mathbf{u},$$

over all unit vectors $\mathbf{u}$. By the optimization principle in Theorem 5.43, the maximum is the largest eigenvalue of $X^T X$ which, assuming $X \neq \mathrm{O}$, is the square of the largest singular value of X. Moreover, the maximizing unit vector is one of the unit eigenvectors for this eigenvalue, i.e., one of the dominant unit singular vectors, so $\mathbf{u} = \pm\mathbf{q}_1$. ■

1.5. ♡ Prove Proposition 8.3.

Solution: Let J denote the $m \times m$ centering matrix (7.5). Then,

$$\begin{aligned} S_Y = \underline{Y}^T\,\underline{Y} = (JY)^T(JY) &= (J\,XW)^T(J\,XW) \\ &= W^T(J\,X)^T(J\,X)W = W^T(\underline{X}^T\,\underline{X})W = W^T S_X W. \end{aligned}$$ ■

2.1. ♡ Verify equation (8.21).

Solution: The second equality is immediate since $S = Y^T Y$. Moreover, according to Exercise 2.8, $\operatorname{tr}(U^T Y^T Y U) = \operatorname{tr}((YU)^T YU)$ is the sum of the squared Euclidean norms of the columns of YU, which are simply $Y\mathbf{u}_j$. Thus

$$\operatorname{tr}(U^T Y^T Y U) = \sum_{j=1}^{k} \|\,Y\mathbf{u}_j\,\|^2 = \sum_{j=1}^{k} \mathbf{u}_j^T\,Y^T\,Y\,\mathbf{u}_j.$$ ■

2.3. ♡ This exercise considers the problem of fitting the best subspace in a general inner product norm $\|\mathbf{x}\|_C = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_C} = \sqrt{\mathbf{x}^T C \mathbf{y}}$, where C is symmetric, positive definite. Given points $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{R}^n$, let $X = (\mathbf{x}_1, \ldots, \mathbf{x}_m)^T$ be the corresponding data matrix. Then, given a subspace $V \subset \mathbb{R}^n$, define the distance and squared energy

$$\operatorname{dist}_C(\mathbf{x}, V) = \min\{\, \|\mathbf{x} - \mathbf{y}\|_C \mid \mathbf{y} \in V \,\}, \qquad E_C(V; \mathbf{x}_1, \ldots, \mathbf{x}_m) = \sum_{i=1}^m \operatorname{dist}_C(\mathbf{x}_i, V)^2.$$

(a) Show that the k-dimensional subspace minimizing E_C is the one spanned by the top k eigenvectors $\mathbf{q}_1, \ldots, \mathbf{q}_k$ of the matrix $S = X^T X C$.

(b) What happens if we minimize over affine subspaces $W = \mathbf{a} + V$? What choice of $\mathbf{a}$ is optimal?

(c) Formulate equivalent optimization principles as was done in Remark 8.12.

Solution: (a) Since $\|\mathbf{x}\|_C = \sqrt{\mathbf{x}^T C \mathbf{x}} = \|C^{1/2}\mathbf{x}\|$, we can set $\mathbf{z}_i = C^{1/2}\mathbf{x}_i$ and $W = C^{1/2} V$, so

$$E_C(V; \mathbf{x}_1, \ldots, \mathbf{x}_m) = E(W; \mathbf{z}_1, \ldots, \mathbf{z}_m).$$

Note that the corresponding data matrix for Z is $Z = X C^{1/2}$. By Theorem 8.9, the best subspace is the space W spanned by the top k eigenvectors of $Z^T Z = C^{1/2} X^T X C^{1/2}$, which we denote by $\mathbf{p}_1, \ldots, \mathbf{p}_k$, so $C^{1/2} X^T X C^{1/2} \mathbf{p}_i = \lambda_i \mathbf{p}_i$. The optimal subspace V is thus $V = C^{-1/2} W$, which is spanned by the vectors $\mathbf{q}_i = C^{-1/2}\mathbf{p}_i$, which are the top k eigenvectors of $S = X^T X C$.

(b) With an offset, so $W = \mathbf{a} + V$, we again have

$$\operatorname{dist}_C(\mathbf{x}, W) = \operatorname{dist}_C(\mathbf{x} - \mathbf{a}, V) = \operatorname{dist}(C^{1/2}\mathbf{x}, C^{1/2}\mathbf{a} + C^{1/2} V),$$

so by Lemma 8.8 an optimal value for $C^{1/2}\mathbf{a}$ is the mean $C^{1/2}\mathbf{a} = \dfrac{1}{m}\displaystyle\sum_{i=1}^m C^{1/2}\mathbf{x}_i$. Multiplying by $C^{-1/2}$, we see that the mean vector $\overline{\mathbf{x}}$ is again an optimal choice for $\mathbf{a}$.

(c) The optimization principles can be formulated by part (a) by using $Z = X C^{1/2}$. The equivalent to (8.20) is

$$\min\left\{ \|X C^{1/2} - X C^{1/2} U U^T\|_F^2 \;\middle|\; U^T U = \mathrm{I} \right\}.$$

Make the change of variables $U = C^{1/2} V$ to find

$$\min\left\{ \|(X - X C V V^T) C^{1/2}\|_F^2 \;\middle|\; V^T C V = \mathrm{I} \right\},$$

which is subject to the more natural orthogonality condition in the C inner product. Notice the $C^{1/2}$ appear at the end inside the Frobenius norm amounts to taking the sum of the squared C norms of each row of the matrix $X - C V V^T$.

The equivalent to (8.22) is

$$\max\left\{ \operatorname{tr}(U^T C^{1/2} X^T X C^{1/2} U) \;\middle|\; U^T U = \mathrm{I} \right\}.$$

Again, the change of variables $U = C^{1/2} V$ yields

$$\max\{\operatorname{tr}(V^T C X^T X C V) \mid V^T C V = \mathrm{I}\}.$$

Recalling that the adjoint of V in the C inner product is $V^* = C^{-1}V^TC$, we can use properties of the trace to write this as

$$\max\left\{ \operatorname{tr}(V^*SVC) \,\middle|\, V^TCV = \mathrm{I} \right\}, \qquad \text{where} \qquad S = X^TXC.$$

3.1. ♡ Generate a plot of the singular values for the rows versus blocks of the image in this section. Which ones decay faster?

Solution: The block-wise singular values decay faster.

Python Notebook: Solution to Exercise 3.1 (.ipynb)

3.5. ♡ In this exercise, you will extend the PCA-based compression algorithm from this section to audio compression. Complete the parts (a) through (c); the notebook below will help you get started.

Python Notebook: Audio Compression (.ipynb)

(a) Use the block-based image compression algorithm described in this section for audio compression. You can use any audio file you like; thePython notebook linked above downloads a classical music sample from the textbook GitHub website. A stereo audio signal is an array of size $n \times 2$, where n is the number of samples. Use blocks of size $N \times 2$ for compression.

(b) Plot the top $k = 10$ or so principal components. They should look suspiciously like sinusoids.

(c) When you play back the compressed audio file, you will likely hear some static noise artifacts, even at very low compression rates. These are caused by blocking artifacts, where the signals do not match up on the edges of the blocks used for compression, which introduces discontinuities into the signal. This is similar to the blocking artifacts we observed in image compression in this section, however, the artifacts are more noticeable in audio than in images.

To fix this, audio compression algorithms use overlapping blocks, and apply a windowing function in order to smoothly patch together the audio in each block. The blocks are structured so that half of the first block overlaps with half of the second block, and so on. To implement this in Python, just shift the signal by half of the block width, and apply the `image_to_patches` function on the original and shifted signals. Then compress and decompress both signals. After decompressing, and before converting back from the block format to the audio signal, you'll need to multiply by a windowing function to smooth the transition between blocks. If the block size is $N \times 2$, then each channel should be multiplied by a window function w_i, $i = 0, 1, \ldots, N-1$. A common window function that is used, for example, in mp3 compression, is

$$w_i = \sin^2\left[\left(i + \frac{1}{2}\right)\frac{\pi}{N}\right].$$

After you decompress and apply the window, undo the shift and add the signals together to get the decompressed audio. Does this improve the audio quality? As a note, in order to make sure the shifted signals add up correctly, we need that

$$w_i + w_{i+N/2} = 1.$$

As an exercise, the reader should check that the window function above satisfies this condition, which is called the *Princen-Bradley condition*.

Solution:

Python Notebook: Audio Compression Solutions (.ipynb)

4.4. ♡ Write Python code to implement the version of LDA where a singular within class covariance is handled according to Exercise 4.2, and compare to covariance shrinkage on MNIST data. *Hint*: Instead of trying to figure out exactly which singular values are zero, a more numerically stable approach is to truncate all singular values less than a threshold $\varepsilon > 0$ to zero. In this problem, it works well to, for example, just take the top 100 eigenvectors of S_w on MNIST.

Solution:

Python Notebook: Solution to Exercise 4.4. (.ipynb)

4.5. ♡ Show that the solution of (8.47) is the matrix $U = Q_k$ whose columns are the top k discriminating directions. *Hint*: See Remark 8.12.

Solution: Let us write $V = S_w^{1/2} U$, so that $U^T S_w U = V^T V$. Then

$$\operatorname{tr}(U^T S_b U) = \operatorname{tr}\left((S_w^{-1/2} V)^T S_b S_w^{-1/2} V\right) = \operatorname{tr}\left(V^T S_w^{-1/2} S_b S_w^{-1/2} V\right).$$

Hence, in terms of the matrix V, the trace LDA problem (8.47) is

$$\max_{V^T V = \mathrm{I}} \left(V^T S_w^{-1/2} S_b S_w^{-1/2} V\right).$$

According to Remark 8.12, the solution is the $m \times k$ matrix $V = (\mathbf{q}_1 \ \ldots \ \mathbf{q}_k)$ whose columns are the top k eigenvectors of the symmetric positive semidefinite matrix $S_w^{-1/2} S_b S_w^{-1/2}$. Then $\mathbf{p}_i = S_w^{-1/2} \mathbf{q}_i$ satisfy $S_b \mathbf{p}_i = S_b S_w^{-1/2} \mathbf{q}_i = \lambda_i S_w^{1/2} \mathbf{q}_i = \lambda_i S_w \mathbf{p}_i$ and so are the LDA discriminating directions. We then set $U = Q_k = (\mathbf{p}_1 \ \ldots \ \mathbf{p}_k) = S_w^{-1/2} V$. ■

5.3. ♡ Let C be a positive definite symmetric matrix and define the distance matrix D_X^C to be the distance matrix in the norm $\|\mathbf{x}\|_C = \sqrt{\mathbf{x}^T C \mathbf{x}}$, with entries $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_C^2$. Generalize Proposition 8.21 and Theorem 8.22 to this setting. In particular, how do you construct the optimal isometric embedding in this case?

Solution: The key is to note that $\|\mathbf{x}\|_C = \|C^{1/2}\mathbf{x}\|$, from which it follows that $D_X^C = D_Y$ where $\mathbf{y}_i = C^{1/2}X$, so $Y = XC^{1/2}$ and $X = YC^{-1/2}$. Thus, (8.51) becomes

$$-\tfrac{1}{2}\, J\, D_X^C J = \underline{X}\, C\, \underline{X}^T.$$

From Theorem 8.22, the requirement is again that $J\, D_X^C J$ is negative semidefinite. The optimal isometric embedding is $X_k = Q_k \Lambda_k^{1/2} C^{-1/2}$, where

$$-\tfrac{1}{2}\, J\, D_X^C J = Q_k \Lambda_k Q_k^T$$

is the reduced spectral decomposition. ■

5.5. ♡ Show that there do not exist three points $z_1, z_2, z_3 \in \mathbb{R}$ that satisfy

$$|z_1 - z_2| = |z_1 - z_3| = |z_2 - z_3| = 1.$$

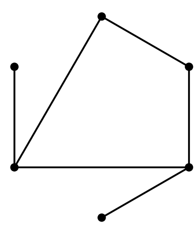
Solution: Let us assume $z_1 < z_2 < z_3$, which would necessitate $z_2 - z_1 = 1$ and $z_3 - z_2 = 1$, and so $z_3 - z_1 = z_3 - z_2 + z_2 - z_1 = 2 \neq 1$. ■

Chapter 9

Graph Theory and Graph-based Learning

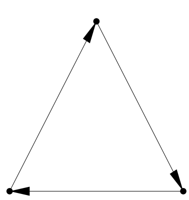
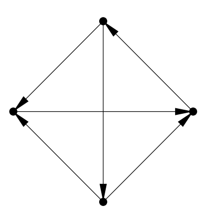
1.1. Sketch the graphs corresponding to the following adjacency matrices.

$$(a)\heartsuit \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}; \quad (b) \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}; \quad (c)\heartsuit \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

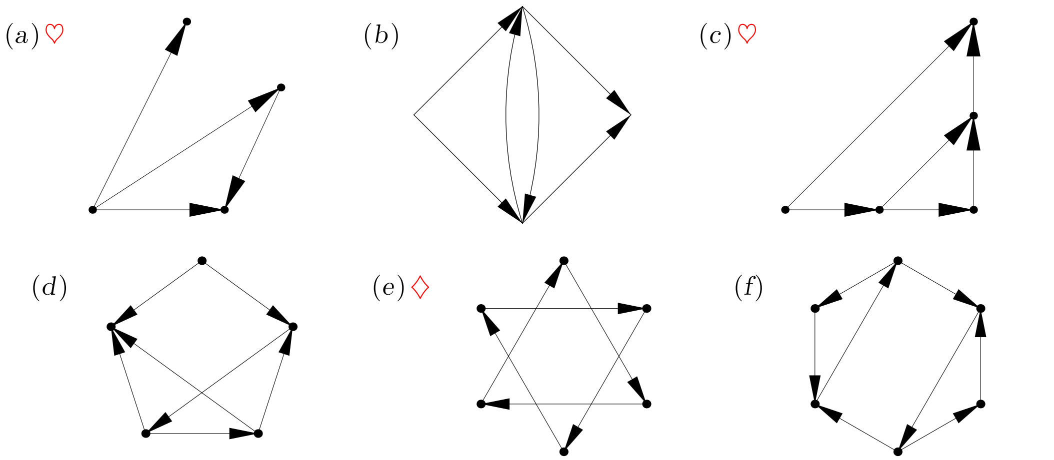
Solution: (a) (c) 

1.2. Sketch the digraphs corresponding to the following adjacency matrices.

$$(a)\heartsuit \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}; \; (b)\diamondsuit \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}; \; (c)\heartsuit \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}; \; (d) \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

Solution: (a)  (c) 

1.3. Write out an adjacency matrix for the following digraphs.



Solution: We label the nodes in order from top to bottom and, when at the same height, from left to right.

$$(a)\ \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \quad (c)\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

1.4. Write out an adjacency matrix for graphs given by the edges of the Platonic solids: (a)♡ tetrahedron, (b)♡ cube, (c)♢ octahedron, (d) dodecahedron, and (e) icosahedron.

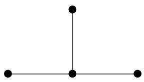
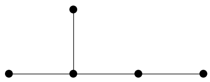
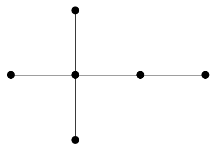
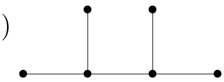
Solution: (a) Tetrahedron: $\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$;

$$(b)\ \text{Cube:}\quad \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{pmatrix}.$$

1.6.♡ *True or false*: Let A be the adjacency matrix for an unweighted digraph. Then the underlying unweighted graph has adjacency matrix $A = \widehat{A} + \widehat{A}^T$.

Solution: True if there is only one edge between any pair of nodes, but false if there are two edges since then the corresponding entry of $\widehat{A} + \widehat{A}^T$ is 2.

1.11. A connected graph is called a *tree* if it has no circuits. (a) Find an adjacency matrix for each of the following trees:

(i) $\heartsuit$  (ii) $\heartsuit$  (iii) $\diamondsuit$  (iv) 

(b) $\diamondsuit$ Draw all distinct trees with 5 nodes, and write down the corresponding adjacency matrices. (c) Prove that any two nodes in a tree are connected by one and only one path.

Solution: (a) (i) $\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$, (ii) $\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$.

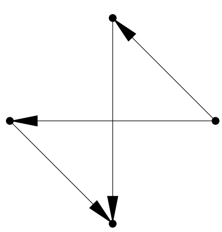
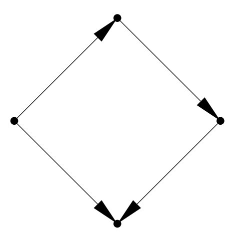
2.2. Verify Euler's formula for each of the Platonic solids of Exercise 1.4.

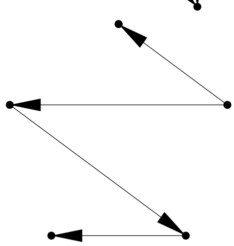
Solution: (a) Tetrahedron: number of nodes $= 4$, number of edges $= 6$, number of circuits $= 3$, Euler's formula: $4 + 3 = 6 + 1$;
(b) Cube: number of nodes $= 8$, number of edges $= 12$, number of circuits $= 5$, Euler's formula: $8 + 5 = 12 + 1$.

2.3. Draw the digraph represented by the following incidence matrices:

(a) $\heartsuit$ $\begin{pmatrix} -1 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}$, (b) $\begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$, (c) $\heartsuit$ $\begin{pmatrix} 0 & 1 & 0 & 0 & -1 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & -1 & 1 & 0 & 0 \end{pmatrix}$,

(d) $\diamondsuit$ $\begin{pmatrix} -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & -1 & 0 & 1 \end{pmatrix}$, (e) $\begin{pmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$.

Solution: (a)  or, equivalently, 

(c) 

2.4. For each of the digraphs in Exercise 1.3, see if you can determine a collection of independent circuits of the underlying graph. Verify your answer by writing out the incidence matrix and constructing a suitable basis of its cokernel.

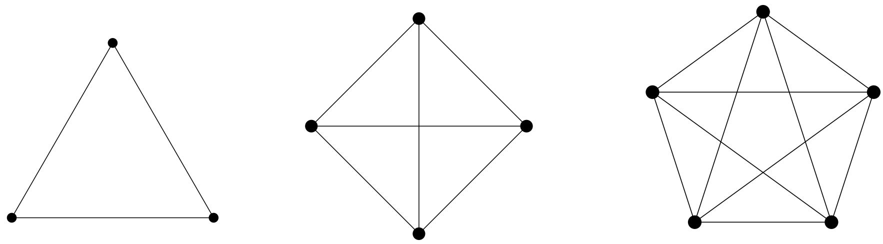
Solution: (a) $\begin{pmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}$, 1 circuit: $\begin{pmatrix} 0 \\ -1 \\ -1 \\ 1 \end{pmatrix}$;

(c) $\begin{pmatrix} -1 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$, 2 circuits: $\begin{pmatrix} -1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1 \\ 1 \\ 0 \\ 1 \end{pmatrix}$.

2.5. ♡ A *complete graph* $\mathcal{G}_m$ on m nodes has one edge joining every distinct pair of nodes. (*a*) Draw $\mathcal{G}_3$, $\mathcal{G}_4$ and $\mathcal{G}_5$. (*b*) Choose an orientation for each edge and write out the resulting incidence matrix of each digraph. (*c*) How many edges does $\mathcal{G}_n$ have? (*d*) How many independent circuits? (*e*) Find a spanning tree and the corresponding basic circuits.

Solution:

(*a*)



(*b*) $\begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}$, $\begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}$, $\begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$.

(*c*) $\frac{1}{2}m(m-1)$; (*d*) $\frac{1}{2}(m-1)(m-2)$; (*e*) One option is to choose all $m-1$ edges starting at a single node, say node 1. The basic circuits are the triangles with nodes $1, i, j$ for $1 < i < j \le m$.

2.11. ♡ *True or false*: If N and $\widetilde{N}$ are incidence matrices of the same size and $\operatorname{coker} N = \operatorname{coker} \widetilde{N}$, then the corresponding digraphs are equivalent.

Solution: False. For example, any two inequivalent trees, cf. Exercise 1.11, with the same number of nodes have incidence matrices of the same size, both with trivial cokernels, so $\operatorname{coker} N = \operatorname{coker} \widetilde{N} = \{\mathbf{0}\}$. As another example, the incidence matrices

$$N = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{pmatrix} \qquad \text{and} \qquad \widetilde{N} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \end{pmatrix}$$

both have cokernel basis $(1,1,1,0,0)^T$, but do not represent equivalent digraphs.

3.2. Determine the graph Laplacian and its spectrum for the graphs with adjacency matrices listed in Exercise 1.1.

Solution: (a) $L = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}$, eigenvalues $= 3, 1, 0$;

$$(c)\ L = \begin{pmatrix} 2 & -1 & 0 & -1 & 0 & 0 \\ -1 & 2 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ -1 & 0 & -1 & 3 & -1 & 0 \\ 0 & -1 & 0 & -1 & 3 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}, \quad \text{eigenvalues} = 4.8136, 3., 2.5293, 1., 0.65708, 0.$$

3.5. ♡ In Proposition 9.21, assume that N is the incidence matrix for a weighted digraph $\widehat{\mathcal{G}}$, without the restriction that each pair of nodes (i,j) has at most one directed edge between them. Show that $L = N^T C N$ is the graph Laplacian for the underlying weighted graph $\mathcal{G}$.

Solution: Referring to the displayed formula in the proof of Proposition 9.21, the weight w_{ij} in the second summation will equal the sum of the weights $w_{i_k j_k}$ of all edges with $i_k = i$ and $j_k = j$ or $i_k = j$ and $j_k = i$, i.e., of all the edges connecting node i and node j in either direction, and hence equals the weight of the edge connecting i and j in $\mathcal{G}$. With this interpretation, the formula continues to identify $N^T C N = L$. ■

3.6. ♡ Let $\mathcal{G}$ be a connected graph with m nodes and with graph Laplacian matrix L. Let $P = (\mathrm{I} \ \ -\mathbf{1})$ be the $(m-1) \times m$ matrix whose first $m-1$ columns form the $(m-1) \times (m-1)$ identity matrix and whose last column has all -1 entries.

(a) Show that the $(m-1) \times (m-1)$ matrix $P L P^T$ is positive definite.

(b) Let $\mathbf{b} \in \mathbb{R}^m$ satisfy $\mathbf{b} \cdot \mathbf{1} = 0$, and let $\mathbf{y} \in \mathbb{R}^m$ be the unique solution of $P L P^T \mathbf{y} = P \mathbf{b}$. Show that $\mathbf{x} = P^T \mathbf{y}$ solves $L\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \cdot \mathbf{1} = 0$.

(c) Suppose $\mathbf{b} \cdot \mathbf{1} \neq 0$ in part (b). What equation does $\mathbf{x} = P^T \mathbf{y}$ satisfy?

Solution: (a) Note that if $\mathbf{y} \in \mathbb{R}^{m-1}$, then

$$\mathbf{x} = P^T \mathbf{y} = \left(y_1, \ldots, y_{m-1}, -y_1 - \cdots - y_{m-1} \right)^T \in \mathbf{1}^\perp$$

lies in the orthogonal complement to the ones vector. Moreover, if $\mathbf{y} \neq \mathbf{0}$, then $\mathbf{x} \neq \mathbf{0}$. Since $\mathcal{G}$

is connected, the kernel of L is spanned by $\mathbf{1}$. Thus, by Corollary 4.9, $\mathbf{y}^T P L P^T \mathbf{y} = \mathbf{x}^T L \mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$.

(b) Part (a) shows $\mathbf{x} \cdot \mathbf{1} = 0$. Moreover, $P(L\mathbf{x} - \mathbf{b}) = \mathbf{0}$, hence $L\mathbf{x} - \mathbf{b} \in \ker P$, which is spanned by $\mathbf{1}$. Together these imply $L\mathbf{x} - \mathbf{b} = \mathbf{0}$.

(c) We have $L\mathbf{x} = \mathbf{b} - \bar{b}\,\mathbf{1}$ where $\bar{b} = \mathbf{b} \cdot \mathbf{1}/m = (b_1 + \cdots + b_m)/m$ is the mean of $\mathbf{b}$. The right hand side is the centered version of $\mathbf{b}$, with mean 0. Indeed, $P^T(L\mathbf{x} - \mathbf{b}) = \mathbf{0}$, and hence $L\mathbf{x} - \mathbf{b} = c\mathbf{1} \in \ker P^T = \ker L$. Thus, $\mathbf{1}^T(L\mathbf{x} - \mathbf{b}) = -\mathbf{1}^T\mathbf{b} = mc$, hence $c = -\mathbf{1}^T\mathbf{b}/m$.

4.1. ♡ Show that the quantity $(m-k)k$ is maximized over integers $0 \leq k \leq m$ by setting $k = \frac{1}{2}m$ when m is even, and $k = \frac{1}{2}(m-1)$ or $k = \frac{1}{2}(m+1)$ when m is odd.

Solution: Complete the square: $(m-k)k = -(k - \frac{1}{2}m)^2 + \frac{1}{4}m^2$. Thus the left hand side is maximized by minimizing $|\,k - \frac{1}{2}m\,|$ over all integers k, which is achieved in the indicated manner.

4.4. ♡ Suppose we allow loops in a graph, by permitting one or more diagonal entries of the weight matrix to be positive: $w_{ii} > 0$. Recall from Exercise 3.3 that the Laplacian matrix is unchanged by loops. Is this also true for the modularity matrix?

Solution: No, because adding in loops at the nodes with weights w_{ii} changes $\widehat{d}_i = d_i + w_{ii}$, and hence, in view of (9.42), the entries of the new modularity matrix $\widehat{M}$ become

$$\widehat{m}_{ij} = w_{ij} - \frac{(d_i + w_{ii})(d_j + w_{jj})}{d_1 + \cdots + d_m + w_{11} + \cdots + w_{mm}},$$

which are clearly not the same as m_{ij}. For a simple example, consider the graph with 2 nodes with weight matrix $W = \begin{pmatrix} w & 1 \\ 1 & 0 \end{pmatrix}$ where $w \geq 0$. Then $\mathbf{d} = (w+1, 1)^T$ and $\mathbf{d}^T\mathbf{1} = w + 2$. Thus, for example, the lower right entry of M is $m_{22} = 0 - \dfrac{d_2 d_2}{\mathbf{d}^T\mathbf{1}} = -\dfrac{1}{w+2}$, which clearly depends on the value of the diagonal entry w.

Intuitively, the reason is that modularity counts how many edges are in a given subset of the graph, and adding self loops affects this count. In contrast, the graph Laplacian is connected to graph cuts, which do not see loops, since you would never cut a loop when clustering a graph.

4.7. ♡ Implement spectral modularity optimization for community detection in Python via the power method and detect communities in Zachary's karate club graph and Krebs' political books graph to reproduce the results from this section. Try other graphs in the `graphlearning` package as well. *Hint*: You will have to use a spectral shift of the modularity matrix, as we did for finding the Fiedler vector with the power iteration; see Exercise 4.5(a).

Solution:

Python Notebook: Solution to Exercise 4.7. (.ipynb)

5.2. ♡ Let $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{R}^n$, and let T be the $m \times m$ matrix with entries $t_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$. Show that T satisfies the triangle inequality $t_{ik} \le t_{ij} + t_{jk}$ for all $1 \le i, j, k \le m$.

Solution: This is immediate from the triangle inequality for the norm:

$$t_{ik} = \|\mathbf{x}_i - \mathbf{x}_k\| = \|(\mathbf{x}_i - \mathbf{x}_j) + (\mathbf{x}_j - \mathbf{x}_k)\| \le \|\mathbf{x}_i - \mathbf{x}_j\| + \|\mathbf{x}_j - \mathbf{x}_k\| = t_{ij} + t_{jk}. \quad ■$$

5.3. ♡ Show that the dynamic programming principle (9.52) can be reformulated as

$$t_i = 0 \quad \text{for} \quad i \in \mathcal{D}, \quad \text{and} \quad \max_j w_{ji}(t_i - t_j) = 1 \quad \text{for} \quad i \in \mathcal{D}^c. \tag{9.56}$$

Solution: Note that we can rewrite (9.52) to read

$$\min_j \, w_{ji}^{-1} \left(w_{ji}(t_j - t_i) + 1 \right) = 0.$$

Indeed, we simply subtract t_i from both sides of (9.52) and factor out w_{ji}^{-1}. Now, the term w_{ji}^{-1} can be removed, while still keeping the minimum value at zero; indeed, this weight is positive so the term in parentheses must vanish at the minimum. Therefore,

$$\min_j \left[w_{ji}(t_j - t_i) + 1 \right] = 0$$

for all nodes i in the graph. We can further rewrite this by multiplying by -1 on both sides to obtain (9.56). ■

5.4. ♡ Let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ be a disconnected graph and $\mathcal{D} \subset \mathcal{N}$. Suppose there is a connected component $\mathcal{H} \subset \mathcal{N}$ for which $\mathcal{H} \cap \mathcal{D} = \varnothing$. Show that there is no solution $\mathbf{t} \in \mathbb{R}^m$ to the dynamic programming principle (9.52). *Hint*: Suppose a solution $\mathbf{t}$ exists, and consider the node i that minimizes t_i over $i \in \mathcal{H}$. It is easier to work with the reformulation (9.56).

Solution: If $i \in \mathcal{H}$ is a point where t_i is minimal over $i \in \mathcal{H}$, then $t_i - t_j \le 0$ for all $j \in \mathcal{H}$, and so $w_{ji}(t_i - t_j) \le 0$ for all j, which violates (9.56). ■

5.5. ♡ Let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ be a connected graph with m nodes, and let $\mathcal{D} \subset \mathcal{N}$. Assume that $\mathbf{u}, \mathbf{v} \in \mathbb{R}^m$ satisfy the reformulated dynamic programming principle (9.56) with inequalities

$$\max_j \, w_{ji}(u_i - u_j) \le 1 \le \max_j \, w_{ji}(v_i - v_j)$$

for all $i \in \mathcal{D}^c$, and suppose that $u_i \le 0 \le v_i$ for $i \in \mathcal{D}$. Show that $\mathbf{u} \le \mathbf{t}_{\mathcal{D}} \le \mathbf{v}$.
Hint: Use a similar argument to the proof of Lemma 9.41.

Solution: Let $\mathbf{t} = \mathbf{t}_{\mathcal{D}}$, and suppose to the contrary that $u_i > t_i$ for some $i \in \mathcal{D}^c$. Then there exists $\lambda > 1$ such that $u_i > \lambda t_i$. Let i be the index where $u_i - \lambda t_i$ is largest. Since the difference is positive, $i \in \mathcal{D}^c$. Therefore $u_i - \lambda t_i \ge u_j - \lambda t_j$ for all $j = 1, \ldots, m$. Rearranging the inequality and multiplying the result by w_{ji} on both sides yields $w_{ji}(u_i - u_j) \ge \lambda w_{ji}(t_i - t_j)$ for all $j = 1, \ldots, m$. Maximizing both sides of the latter inequality over j implies that $\lambda \le 1$, which is a contradiction. The proof that $\mathbf{t}_{\mathcal{D}} \le \mathbf{v}$ is similar. ■

6.1. ♡ Let $\mathbf{x}_k$ satisfy the diffusion equation (9.60) and set $\mathbf{z}_k = D^{-1}\mathbf{x}_k$. Show that $\mathbf{z}_k$ solves the adjoint equation (9.62).

Solution: Since $W = W^T$ is symmetric, so is the graph Laplacian $L^T = L$. Thus

$$\begin{aligned}\mathbf{z}_{k+1} = D^{-1}\mathbf{x}_{k+1} &= D^{-1}\mathbf{x}_k - D^{-1}L_{\mathrm{rw}}^T\mathbf{x}_k \\ &= D^{-1}\mathbf{x}_k - D^{-1}L^T D^{-1}\mathbf{x}_k = \mathbf{z}_k - D^{-1}L\,\mathbf{z}_k = \mathbf{z}_k - L_{\mathrm{rw}}\mathbf{z}_k.\end{aligned}$$ ■

6.3. ♡ The symmetric normalized graph Laplacian (9.65) defines another type of diffusion on graphs, which is given by

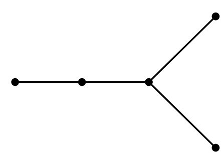
$$\mathbf{y}_{k+1} = \mathbf{y}_k - L_{\mathrm{sym}}\mathbf{y}_k. \tag{9.82}$$

Let $\mathbf{x}_k$ satisfy the diffusion equation (9.60) and set $\mathbf{y}_k = D^{-1/2}\mathbf{x}_k$. Show that $\mathbf{y}_k$ solves the symmetric diffusion equation (9.82).

Solution:

$$\begin{aligned}\mathbf{y}_{k+1} = D^{-1/2}\mathbf{x}_{k+1} &= D^{-1/2}\mathbf{x}_k - D^{-1/2}L_{\mathrm{rw}}^T\mathbf{x}_k = \mathbf{y}_k - D^{-1/2}L\,D^{-1}\mathbf{x}_k \\ &= \mathbf{y}_k - \left(D^{-1/2}L\,D^{-1/2}\right)D^{-1/2}\mathbf{x}_k = \mathbf{y}_k - L_{\mathrm{sym}}\mathbf{y}_k.\end{aligned}$$ ■

6.4. ♡ Let $\mathbf{x}_k$ satisfy the diffusion equation (9.60). Give an example of a graph where $\mathbf{x}_0 = \mathbf{1}$, but $\mathbf{x}_k \neq \mathbf{1}$ when $k \geq 1$.

Solution: An example is  In this case,

$$D = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \qquad L = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{pmatrix},$$

and so $L_{\mathrm{rw}}^T = \begin{pmatrix} 1 & -\frac12 & 0 & 0 & 0 \\ -1 & 1 & -\frac12 & 0 & 0 \\ 0 & -\frac12 & 1 & -1 & -1 \\ 0 & 0 & -\frac13 & 1 & 0 \\ 0 & 0 & -\frac13 & 0 & 1 \end{pmatrix}$. Here are the first few iterates:

$$\mathbf{x}_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{x}_1 = \begin{pmatrix} \frac12 \\ \frac43 \\ \frac52 \\ \frac13 \\ \frac13 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} \frac23 \\ \frac43 \\ \frac43 \\ \frac56 \\ \frac56 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} \frac23 \\ \frac{1}{0}9 \\ \frac73 \\ \frac49 \\ \frac49 \end{pmatrix}, \quad \mathbf{x}_4 = \begin{pmatrix} \frac59 \\ \frac{13}{9} \\ \frac{13}{9} \\ \frac79 \\ \frac79 \end{pmatrix}, \quad \mathbf{x}_5 = \begin{pmatrix} \frac{13}{18} \\ \frac{28}{27} \\ \frac{41}{18} \\ \frac{13}{27} \\ \frac{13}{27} \end{pmatrix}.$$

The graph is not aperiodic, and the diffusion process has a periodic limiting behavior, switching back and forth between $\left(\frac34, 1, \frac94, \frac12, \frac12\right)^T$ and $\left(\frac12, \frac32, \frac32, \frac34, \frac34\right)^T$.

6.5. ♡ Show that the solution $\mathbf{x}_k$ of the diffusion equation (9.60) is given by

$$\mathbf{x}_k = D \sum_{i=1}^{m} (1-\lambda_i)^k \, (\mathbf{x}_0 \cdot \mathbf{p}_i)\, \mathbf{p}_i, \tag{9.83}$$

where $\mathbf{p}_1, \ldots, \mathbf{p}_m$ are the orthonormal eigenvector basis of L_{rw} with eigenvalues $\lambda_1, \ldots, \lambda_m$.

Solution: Using Exercise 6.1 and (9.68),

$$\mathbf{x}_k = D\,\mathbf{z}_k = D \sum_{i=1}^{m} (1-\lambda_i)^k \langle\, \mathbf{p}_i \,, \mathbf{z}_0 \,\rangle_D \, \mathbf{p}_i .$$

Furthermore, $\langle\, \mathbf{p}_i \,, \mathbf{z}_0 \,\rangle_D = \mathbf{p}_i^T D\,\mathbf{z}_0 = \mathbf{p}_i^T \mathbf{x}_0 = \mathbf{x}_0 \cdot \mathbf{p}_i$. Substituting this into the preceding equation produces the result. ■

6.7. ♡ State and prove Theorem 9.61 for $\mathbf{x}_k$ and $\mathbf{y}_k$, the respective solutions of the diffusion equations (9.60) and (9.82).

Solution: According to Exercises 6.1 and 6.3, $\mathbf{x}_k = D\,\mathbf{z}_k$ and $\mathbf{y}_k = D^{1/2}\,\mathbf{z}_k$. Thus, the limiting distributions are

$$\mathbf{x}_\infty = D \sum_{\mathcal{H}} \frac{\mathbf{1}_{\mathcal{H}} \cdot \mathbf{x}_0}{\|\,\mathbf{1}_{\mathcal{H}}\,\|_D^2}\, \mathbf{1}_{\mathcal{H}}, \qquad \mathbf{y}_\infty = D^{1/2} \sum_{\mathcal{H}} \frac{(D^{1/2}\,\mathbf{1}_{\mathcal{H}}) \cdot \mathbf{y}_0}{\|\,\mathbf{1}_{\mathcal{H}}\,\|_D^2}\, \mathbf{1}_{\mathcal{H}}.$$

Moreover,

$$\|\,\mathbf{x}_k - \mathbf{x}_\infty\,\|_{D^{-1}} \le \tau^k \,\|\,\mathbf{x}_0\,\|_{D^{-1}}, \qquad \|\,\mathbf{y}_k - \mathbf{y}_\infty\,\| \le \tau^k \,\|\,\mathbf{y}_0\,\|, \qquad k \ge 0,$$

where $\tau = \min\left\{\,|\,1-\lambda\,|,\ |\,1-\lambda_m\,|\,\right\}$. In particular, if $\mathcal{G}$ is aperiodic, then $\tau < 1$ and hence $\mathbf{x}_k \to \mathbf{x}_\infty$ and $\mathbf{y}_k \to \mathbf{y}_\infty$ as $k \to \infty$. ■

6.10. ♡ Let $\mathbf{x}_\infty$ solve (9.80) with W not necessarily symmetric. Let $\mathbf{v}$ be a probability vector. Show that $\|\,\mathbf{x}_\infty - \mathbf{v}\,\|_1 \le 2\,\alpha$.

Solution:

$$\|\,\mathbf{x}_\infty - \mathbf{v}\,\|_1 = \|\,\alpha\, P\,\mathbf{x}_\infty - \alpha\,\mathbf{v}\,\|_1 \le \alpha\,\|\,P\,\mathbf{x}_\infty\,\|_1 + \alpha\,\|\,\mathbf{v}\,\|_1 \le \alpha\left(\,\|\,\mathbf{x}_\infty\,\|_1 + \alpha\,\|\,\mathbf{v}\,\|_1\,\right) = 2\,\alpha,$$

because both $\mathbf{v}$ and $\mathbf{x}_\infty$ are probability vectors. ■

6.15. ♡ Let $\mathcal{G}$ be a connected weighted graph. Show that there exists $C > 0$, depending only on $\mathcal{G}$, such that

$$\sum_{i=1}^{m} u_i^2 \le \frac{C}{2} \sum_{i,j=1}^{m} w_{ij}(u_i - u_j)^2, \tag{9.86}$$

for all $\mathbf{u} \in \mathbb{R}^m$ satisfying $\mathbf{u} \cdot \mathbf{1} = 0$. The result is known as a *Poincaré inequality*.

Solution: By Proposition 9.16, the inequality (9.86) is equivalent to

$$\|\,\mathbf{u}\,\|^2 \le C\,\mathbf{u}^T L\,\mathbf{u} \qquad \text{or} \qquad \frac{\mathbf{u}^T L\,\mathbf{u}}{\|\,\mathbf{u}\,\|^2} \ge C^{-1},$$

for all $\mathbf{u} \in \mathbb{R}^m$ such that $\mathbf{u} \cdot \mathbf{1} = 0$. Since $\mathbf{1} \in \ker L$, which is one-dimensional, the result follows from the optimization principle in Theorem 5.47, upon setting $C = \lambda_2^{-1} > 0$. ■

7.1. ♡ Prove that the diffusion distance satisfies the triangle inequality (9.88).

Solution: By the triangle inequality for the C norm:

$$d_{i\ell}^{(k)} = \| P^k(\mathbf{e}_i - \mathbf{e}_\ell) \|_C = \| P^k(\mathbf{e}_i - \mathbf{e}_j) + P^k(\mathbf{e}_j - \mathbf{e}_\ell \|_C \\ \leq \| P^k(\mathbf{e}_i - \mathbf{e}_j) \|_C + \| P^k(\mathbf{e}_j - \mathbf{e}_\ell) \|_C = d_{ij}^{(k)} + d_{j\ell}^{(k)}.$$ ■

7.4. ♡ Let S be a self-adjoint matrix for the inner product $\langle \mathbf{x}, \mathbf{y} \rangle_C = \mathbf{x}^T C \mathbf{y}$, where C is symmetric positive definite, and let $S = Q \Lambda Q^T C$ be its spectral decomposition. Follow the proof of Theorem 9.69 to show that $\| S^k(\mathbf{e}_i - \mathbf{e}_j) \|_C = \| \mathbf{x}_i - \mathbf{x}_j \|$, where $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{R}^m$ are the rows of the matrix $X = C Q \Lambda^k$.

Solution: Since $S^k = Q \Lambda^k Q^T C$ and $Q^T C Q = \mathrm{I}$, we have

$$\| S^k(\mathbf{e}_i - \mathbf{e}_j) \|_C^2 = (\mathbf{e}_i - \mathbf{e}_j)^T C Q \Lambda^k Q^T C Q \Lambda^k Q^T C (\mathbf{e}_i - \mathbf{e}_j) \\ = (\mathbf{e}_i - \mathbf{e}_j)^T X X^T (\mathbf{e}_i - \mathbf{e}_j) = \| X^T(\mathbf{e}_i - \mathbf{e}_j) \|^2 = \| \mathbf{x}_i - \mathbf{x}_j \|^2.$$ ■

7.6. ♡ Use Remark 8.12 to show that the first k eigenvectors $\mathbf{q}_1, \ldots, \mathbf{q}_k$ of the graph Laplacian L solve the optimization problem

$$\min \sum_{i=1}^{k} \mathbf{u}_i^T L \mathbf{u}_i = \mathrm{tr}\,(U^T L U), \tag{9.97}$$

where the minimum is over all $m \times k$ matrices $U = (\,\mathbf{u}_1 \ \ldots \ \mathbf{u}_k\,)$ with orthonormal columns. What is the minimum value of this optimization problem?

Solution: This is immediately deduced from Remark 8.12 and Theorem 8.9, by replacing $Y^T Y$ with $-L$. The minimum value is the sum $\lambda_1 + \cdots + \lambda_k$ of the first k eigenvalues of L.

8.2. ♡ Show that equation (9.106) holds.

Solution: Note first that, using the formula for γ in (9.100),

$$\nabla_{\mathbf{z}_\ell} \log \left(\sum_{\substack{i,j=1 \\ i \neq j}}^{m} \left(1 + \| \mathbf{z}_i - \mathbf{z}_j \|^2 \right)^{-1} \right) = \frac{1}{\gamma} \sum_{\substack{i,j=1 \\ i \neq j}}^{m} \nabla_{\mathbf{z}_\ell} \left[\frac{1}{1 + \| \mathbf{z}_i - \mathbf{z}_j \|^2} \right] \\ = \frac{2}{\gamma} \sum_{j=1}^{m} \frac{-2\,(\mathbf{z}_\ell - \mathbf{z}_j)}{\left(1 + \| \mathbf{z}_\ell - \mathbf{z}_j \|^2 \right)^2} = -4\gamma \sum_{j=1}^{m} q_{\ell j}^2 \,(\mathbf{z}_\ell - \mathbf{z}_j).$$ ■

8.4. ♡ Consider a probability vector $\mathbf{p}_\lambda = (\,\lambda, 1 - \lambda\,)^T \in \mathbb{R}^2$, where $0 < \lambda < 1$, and let $h(\lambda)$ be its entropy. (a) Write down a formula for $h(\lambda)$. (b) Show that $h(\lambda)$ is increasing for $0 < \lambda < \frac{1}{2}$ and decreasing for $\frac{1}{2} < \lambda < 1$.

Solution: (a) $h(\lambda) = -\lambda \log \lambda - (1 - \lambda) \log(1 - \lambda)$.

(b) $h'(\lambda) = -\log \lambda + \log(1 - \lambda) = \log(\lambda^{-1} - 1)$, which is positive for $0 < \lambda < \frac{1}{2}$ and negative for $\frac{1}{2} < \lambda < 1$.

8.7. ♡ Suppose we generalize the definition of Q in (9.100) to read

$$q_{ij} = \begin{cases} \dfrac{1}{\gamma}\, f\big(\,\|\mathbf{z}_i - \mathbf{z}_j\|^2\,\big), & i \neq j \\ 0 & i = j, \end{cases} \qquad \text{where} \qquad \gamma = \sum_{\substack{i,j=1 \\ i\neq j}}^{m} f\big(\,\|\mathbf{z}_i - \mathbf{z}_j\|^2\,\big), \tag{9.112}$$

where $f\colon \mathbb{R} \to \mathbb{R}$. Show that the gradient of the corresponding Kullback–Leibler divergence (9.101) is given by

$$\nabla_{\mathbf{z}_i} E = \frac{4}{\gamma} \sum_{\substack{i,j=1 \\ i\neq j}} \left(\frac{p_{ij}}{q_{ij}} - 1 \right) g\big(\,\|\,\mathbf{z}_i - \mathbf{z}_j\,\|^2\,\big)\,(\mathbf{z}_i - \mathbf{z}_j), \tag{9.113}$$

where $g(t) = -\,f'(t)$. Verify that this gives the correct t-SNE gradient when $f(t) = 1/(1+t)$.

Solution: In this case, (9.102) becomes

$$\widehat{E}(\mathbf{z}_1, \ldots, \mathbf{z}_m) = -\sum_{i,j=1}^{m} p_{ij} \log q_{ij} - \sum_{i,j=1}^{m} p_{ij} \log f\big(\,\|\,\mathbf{z}_i - \mathbf{z}_j\,\|^2\,\big) - \log \left(\sum_{\substack{i,j=1 \\ i\neq j}}^{m} f\big(\,\|\mathbf{z}_i - \mathbf{z}_j\|^2\,\big) \right),$$

and hence, using (9.104) and mimicking the computations in (9.105) and (9.106),

$$\begin{aligned} \nabla_{\mathbf{z}_i} E = \nabla_{\mathbf{z}_i} \widehat{E} &= -\sum_{i\neq j}^{m} p_{ij} \frac{f'\big(\,\|\mathbf{z}_i - \mathbf{z}_j\|^2\,\big)}{f\big(\,\|\mathbf{z}_i - \mathbf{z}_j\|^2\,\big)}\, \nabla_{\mathbf{z}_\ell} \|\,\mathbf{z}_i - \mathbf{z}_j\,\|^2 - \frac{4}{\gamma} \sum_{j=1}^{m} f'\big(\,\|\mathbf{z}_i - \mathbf{z}_j\|^2\,\big)(\mathbf{z}_\ell - \mathbf{z}_j) \\ &= \frac{4}{\gamma} \sum_{\substack{\ell,j=1 \\ j\neq \ell}} \left(\frac{p_{\ell j}}{q_{\ell j}} - 1 \right) g\big(\,\|\mathbf{z}_\ell - \mathbf{z}_j\|^2\,\big)\,(\mathbf{z}_\ell - \mathbf{z}_j). \end{aligned}$$

Replacing ℓ by i completes the demonstration.

If $f(t) = 1/(1+t)$, then $g(t) = 1/(1+t)^2$, so $g\big(\,\|\,\mathbf{z}_i - \mathbf{z}_j\,\|^2\,\big) = \gamma^2\, q_{ij}^2$ for $i \neq j$, as in (9.100), and hence (9.113) reduces to (9.107).

8.9. ♡ Implement the SNE algorithm in Python and compare the results to t-SNE on small toy data sets. The Python notebook from this section will be helpful.

Solution:

Python Notebook: Solution to Exercise 8.9. (.ipynb)

8.10.♡ This exercise explores the uniform manifold approximation (UMAP) algorithm [160], which is a variant of t-SNE that can often give better visualizations. The algorithm starts with a weight matrix W constructed over the high dimensional data for which the maximum entry in every row is exactly one; that is $\max_j w_{ij} = 1$ for all i.

(*a*) The UMAP algorithm uses the symmetrization $p_{ij} = w_{ij} + w_{ji} - w_{ij}w_{ji}$, which in particular is not normalized to be a probability matrix. Show that $0 \le p_{ij} \le 1$.

(*b*) For the embedded data points, UMAP uses the weights

$$q_{ij} = \frac{1}{1 + a\,\|\,\mathbf{z}_i - \mathbf{z}_j\,\|^{2b}}, \tag{9.114}$$

for parameters a, b, which are also not normalized. The loss function for UMAP is defined by

$$E(\mathbf{z}_1, \ldots, \mathbf{z}_m) = \sum_{i=1}^{m} \sum_{\substack{j=1 \\ j \neq i}}^{m} \left[p_{ij} \log\left(\frac{p_{ij}}{q_{ij}}\right) + (1 - p_{ij}) \log\left(\frac{1 - p_{ij}}{1 - q_{ij}}\right) \right], \tag{9.115}$$

which is referred to as *fuzzy cross-entropy*. Take $a = b = 1$, and show that

$$\nabla_{\mathbf{z}_i} E = 4 \sum_{j=1}^{m} (1 - q_{ij})^{-1} q_{ij} (p_{ij} - q_{ij})(\mathbf{z}_i - \mathbf{z}_j). \tag{9.116}$$

Solution: (a) Note that $p_{ij} = 1 - (1 - w_{ij})(1 - w_{ji}) \in [0, 1]$.

(*b*) For the purposes of computing the gradient, we may write

$$E = -\sum_{i=1}^{m} \sum_{\substack{j=1 \\ j \neq i}}^{m} \left[\, p_{ij} \log(q_{ij}) + (1 - p_{ij}) \log(1 - q_{ij}) \,\right].$$

Since

$$\nabla_{\mathbf{z}_\ell} q_{ij} = \begin{cases} -2q_{ij}^2(\mathbf{z}_i - \mathbf{z}_j), & \text{if } \ell = i \neq j \\ -2q_{ij}^2(\mathbf{z}_j - \mathbf{z}_i), & \text{if } \ell = j \neq i \\ 0, & \text{if } \ell \neq i, j, \end{cases}$$

we have

$$\begin{aligned} \nabla_{\mathbf{z}_\ell} E &= -\sum_{i=1}^{m} \sum_{\substack{j=1 \\ j \neq i}}^{m} \left[\, p_{ij} q_{ij}^{-1} - (1 - p_{ij})(1 - q_{ij})^{-1} \,\right] \nabla_{\mathbf{z}_\ell} q_{ij} \\ &= 2 \sum_{\substack{j=1 \\ j \neq \ell}}^{m} \left[\, p_{\ell j} q_{\ell j}^{-1} - (1 - p_{\ell j})(1 - q_{\ell j})^{-1} \,\right] q_{\ell j}^2 (\mathbf{z}_\ell - \mathbf{z}_j) \\ &\qquad\qquad + 2 \sum_{\substack{i=1 \\ i \neq \ell}}^{m} \left[\, p_{i\ell} q_{i\ell}^{-1} - (1 - p_{i\ell})(1 - q_{i\ell})^{-1} \,\right] q_{i\ell}^2 (\mathbf{z}_\ell - \mathbf{z}_i) \\ &= 4 \sum_{j=1}^{m} (1 - q_{\ell j})^{-1} q_{\ell j} (p_{\ell j} - q_{\ell j})\, (\mathbf{z}_\ell - \mathbf{z}_j). \end{aligned}$$

9.5. ♡ Implement the following binary semi-supervised learning algorithm in Python. Let $\mathbf{y} \in \mathbb{R}^m$ be the label vector encoding the given labels, with $y_i = +1$ if i is in one class, $y_i = -1$ if i is in the other class, and $y_i = 0$ otherwise. Assume the number of labels in each class is the same. Start from the zero vector $\mathbf{u}_0 = \mathbf{0}$, and run the iterations

$$\mathbf{u}_{k+1} = D^{-1} W \mathbf{u}_k + D^{-1} \mathbf{y} \tag{9.135}$$

until convergence. This is essentially the power iteration for spectral clustering that we say previously in Exercise 4.3, except we are adding the labels as a *source term*. After convergence, threshold the vector $\mathbf{u}_k$ at zero to produce the two classes. Try your algorithm on some simple binary classification problems, like two moons, circles, and pairs of MNIST digits.

Solution:

Python Notebook: Solution to Exercise 9.5. (.ipynb)

9.6. ♡ Under the assumptions of Exercise 9.5, show that $\lim_{k\to\infty} \mathbf{u}_k = \mathbf{u}$, where $\mathbf{u}$ is the unique solution of $L\mathbf{u} = \mathbf{y}$ that satisfies $\langle \mathbf{u}, \mathbf{1} \rangle_D = 0$.

Solution: First, since there are equal numbers of points in both classes, we have $\mathbf{1} \cdot \mathbf{y} = 0$, so $\mathbf{y} \in (\ker L)^\perp = \operatorname{img} L$. This implies that the equation $L\mathbf{u} = \mathbf{y}$ has a solution $\mathbf{u}$ which is unique up to the addition of a term $c\mathbf{1} \in \ker L$ for some $c \in \mathbb{R}$. We choose c so that $\langle \mathbf{1}, \mathbf{u} \rangle_D = 0$. Note that the resulting $\mathbf{u}$ satisfies $\mathbf{u} = D^{-1}W\mathbf{u} + D^{-1}\mathbf{y}$. Thus, $\mathbf{z}_k = \mathbf{u}_k - \mathbf{u}$ satisfies the adjoint diffusion equation $\mathbf{z}_{k+1} = D^{-1}W\mathbf{z}_k$. Since $\mathbf{u}_0 = \mathbf{0}$, we have $\langle \mathbf{1}, \mathbf{z}_0 \rangle = 0$. Then Theorem 9.61 tells us that $\mathbf{z}_k \to \mathbf{0}$ as $k \to \infty$. ■

9.7. ♡ Under the assumptions of Exercises 9.5, 9.6, show that $\mathbf{u}$ is the unique minimizer of $E(\mathbf{u}) = \frac{1}{2}\mathbf{x}^T L \mathbf{x} - \mathbf{y}^T \mathbf{u}$ satisfying $\langle \mathbf{u}, \mathbf{1} \rangle_D = 0$. *Hint*: Use Theorem 9.61.

Solution: Since $\mathbf{y} \cdot \mathbf{1} = 0$ we have $\mathbf{y} \in (\ker L)^T = \operatorname{img} L$, and so, by Theorem 6.9, E admits a minimizer, which solves $L\mathbf{u} = \mathbf{y}$. As in Exercise 9.6, we select the unique solution to this system satisfying the mean zero constraint.

9.8. ♡ For very large sparse graphs, direct methods struggle to solve the Laplacian regularized learning problem (9.120). Implement a spectrally truncated solver, as outlined in Exercise 3.7 in Chapter 5, and apply it to graph-based semi-supervised binary classification on the MNIST data set.

Solution:

Python Notebook: Solution to Exercise 9.8. (.ipynb)

9.12. ♡ Let $\mathcal{G}$ be an unweighted graph with adjacency matrix A. Let $k \geq 0$ such that for each pair of adjacent nodes i, j — so that $a_{ij} = 1$ — there are at least k other nodes that are adjacent to both i and j; that is the number of indices $\ell \neq i, j$ for which $a_{i\ell} = 1 = a_{j\ell}$ is at least k. Show that

$$|\, u_i - u_j \,| \leq \sqrt{\frac{2\,\mathbf{u}^T L\,\mathbf{u}}{k+2}} \tag{9.136}$$

holds for all $\mathbf{u} \in \mathbb{R}^m$ and all adjacent nodes i, j. *Note*: The inequality (9.136) allows us to estimate the difference in the label predictions between two adjacent nodes in terms of the size of the regularization term $\mathbf{u}^T L\,\mathbf{u}$.

Solution: Fix any adjacent pair i, j, and let ℓ be any other index. Then, by the triangle inequality and Cauchy's inequality (6.98),

$$\begin{aligned} |\, u_i - u_j \,|^2 &\leq \big(\,|\, u_i - u_\ell \,| + |\, u_\ell - u_j \,|\,\big)^2 \\ &= (u_i - u_\ell)^2 + (u_j - u_\ell)^2 + 2\,|\, u_i - u_\ell \,|\,|\, u_\ell - u_j \,| \leq 2\big[\,(u_i - u_\ell)^2 + (u_j - u_\ell)^2\,\big]. \end{aligned}$$

Let $J \subset \{1, \ldots, m\}$ be the set of indices corresponding to the adjacent pair i, j, so, by the assumption, $\# J \geq k$. By Proposition 9.16 and the computation above,

$$\begin{aligned} \mathbf{u}^T L \mathbf{u} &\geq (u_i - u_j)^2 + \sum_{\ell \in J}^{m} \big[\, a_{i\ell}(u_i - u_\ell)^2 + a_{j\ell}(u_j - u_\ell)^2 \,\big] \\ &\geq (u_i - u_j)^2 + \sum_{\ell \in J} \big[\,(u_i - u_\ell)^2 + (u_j - u_\ell)^2\,\big] \geq (u_i - u_j)^2 + \frac{1}{2}\sum_{\ell \in J}(u_i - u_j)^2 \\ &\geq \left(1 + \frac{k}{2}\right)(u_i - u_j)^2 = \frac{k+2}{2}\,(u_i - u_j)^2. \end{aligned}$$

■

10.1. ♡ Find a formula for the the reciprocal $1/z$ of a nonzero complex number $z \neq 0$ in terms of its real and imaginary parts. Then write out the formula for complex division z/w. *Hint*: Use the formula $z\,\overline{z} = |\, z \,|^2$.

Solution:

$$\frac{1}{z} = \frac{\overline{z}}{|\, z \,|^2}\,, \qquad z \neq 0, \qquad \text{or, equivalently,} \qquad \frac{1}{x + \mathrm{i}\,y} = \frac{x - \mathrm{i}\,y}{x^2 + y^2}\,.$$

The general formula for complex division is

$$\frac{w}{z} = \frac{w\,\overline{z}}{|\, z \,|^2} \qquad \text{or} \qquad \frac{u + \mathrm{i}\,v}{x + \mathrm{i}\,y} = \frac{(x\,u + y\,v) + \mathrm{i}\,(x\,v - y\,u)}{x^2 + y^2}\,.$$

10.3. ♡ Use the complex orthogonality formulas (9.179) to prove that the real vectors $\mathbf{x}_k, \mathbf{y}_k$ in (9.141) are mutually orthogonal, and, moreover, the Euclidean norm formulas (9.144) hold. *Remark*: A purely real proof, based on trigonometric identities, is doable, but more challenging.

Solution: Using (9.178),

$$1 = \mathbf{z}_k \cdot \mathbf{z}_k = \mathbf{z}_k^T \,\overline{\mathbf{z}}_k = \frac{1}{m}\,(\mathbf{x}_k + \mathrm{i}\,\mathbf{y}_k)^T(\mathbf{x}_k - \mathrm{i}\,\mathbf{y}_k) = \frac{1}{m}\left(\|\,\mathbf{x}_k\,\|^2 + \|\,\mathbf{y}_k\,\|^2\right),$$

and hence $\|\,\mathbf{x}_k\,\|^2 + \|\,\mathbf{y}_k\,\|^2 = m$. In particular, since $\mathbf{y}_0 = \mathbf{0} = \mathbf{y}_n$, the latter holding only when $m = 2\,n$ is even, we have $\|\,\mathbf{x}_0\,\| = \|\,\mathbf{x}_n\,\| = \sqrt{m}$ under the same condition on m. Furthermore, if $1 \le k < \frac{1}{2}\,m$, we have

$$0 = \mathbf{z}_k \cdot \mathbf{z}_{-k} = \mathbf{z}_k^T\,\mathbf{z}_k = \frac{1}{m}\,(\mathbf{x}_k + \mathrm{i}\,\mathbf{y}_k)^T(\mathbf{x}_k + \mathrm{i}\,\mathbf{y}_k) = \frac{1}{m}\left(\|\,\mathbf{x}_k\,\|^2 - \|\,\mathbf{y}_k\,\|^2\right) + \frac{2}{m}\,(\mathbf{x}_k \cdot \mathbf{y}_k)\,\mathrm{i}\,,$$

which, by the previous result, implies $\|\,\mathbf{x}_k\,\|^2 = \|\,\mathbf{y}_k\,\|^2 = \frac{1}{2}\,m$, and, moreover, $\mathbf{x}_k \cdot \mathbf{y}_k = 0$. ■

10.7. ♡ Prove the unnormalized DFT convolution formulas in (9.211).

Solution: In view of (9.184), (9.185), the formulas in (9.210) become

$$G_m(\mathbf{z} * \mathbf{w}) = \sqrt{m}\;F_m(\mathbf{z} * \mathbf{w}) = m\big(\,F_m\,\mathbf{z} \circ F_m\,\mathbf{w}\,\big) = G_m\,\mathbf{z} \circ G_m\,\mathbf{w}.$$

10.8. Use the fast Fourier transform to find the discrete Fourier coefficients for the following functions using the indicated number of sample points. Carefully indicate each step in your analysis.

(a) ♡ $\dfrac{x}{\pi}$, $\;n = 4$; $\quad (b)$ ♢ $\sin x$, $\;n = 8$; $\quad (c)$ ♡ $|\,x - \pi\,|$, $\;n = 8$; $\quad (d)$ $\operatorname{sign}(x - \pi)$, $\;n = 16$.

Solution: (a) $\mathbf{f} = \begin{pmatrix} 0 \\ \frac12 \\ 1 \\ \frac32 \end{pmatrix}, \quad c^{(0)} = \begin{pmatrix} 0 \\ 1 \\ \frac12 \\ \frac32 \end{pmatrix}, \quad c^{(1)} = \begin{pmatrix} \frac12 \\ -\frac12 \\ 1 \\ -\frac12 \end{pmatrix}, \quad c = c^{(2)} = \begin{pmatrix} \frac34 \\ -\frac14 + \frac14\,\mathrm{i} \\ -\frac14 \\ -\frac14 + \frac14\,\mathrm{i} \end{pmatrix};$

$$(c)\;\; \mathbf{f} = \begin{pmatrix} \pi \\ \frac34\,\pi \\ \frac12\,\pi \\ \frac14\,\pi \\ 0 \\ \frac14\,\pi \\ \frac12\,\pi \\ \frac34\,\pi \end{pmatrix}, \; c^{(0)} = \begin{pmatrix} \pi \\ 0 \\ \frac12\,\pi \\ \frac12\,\pi \\ \frac34\,\pi \\ \frac14\,\pi \\ \frac14\,\pi \\ \frac34\,\pi \end{pmatrix}, \; c^{(1)} = \begin{pmatrix} \frac12\,\pi \\ \frac12\,\pi \\ \frac12\,\pi \\ 0 \\ \frac12\,\pi \\ \frac14\,\pi \\ \frac12\,\pi \\ -\frac14\,\pi \end{pmatrix}, \; c^{(2)} = \begin{pmatrix} \frac12\,\pi \\ \frac14\,\pi \\ 0 \\ \frac14\,\pi \\ \frac12\,\pi \\ \frac{1+\mathrm{i}}{8}\,\pi \\ 0 \\ \frac{1-\mathrm{i}}{8}\,\pi \end{pmatrix}, \; c = c^{(3)} = \begin{pmatrix} \frac12\,\pi \\ \frac{\sqrt2+1}{8\sqrt2}\,\pi \\ 0 \\ \frac{\sqrt2-1}{8\sqrt2}\,\pi \\ 0 \\ \frac{\sqrt2-1}{8\sqrt2}\,\pi \\ 0 \\ \frac{\sqrt2+1}{8\sqrt2}\,\pi \end{pmatrix}.$$

10.9. Use the inverse fast Fourier transform to reassemble the sampled function data corresponding to the following discrete Fourier coefficients. Carefully indicate each step.
(a) ♡ $c_0 = c_2 = 1,\ \ c_1 = c_3 = -1,$ (b) ♢ $c_0 = c_1 = c_4 = 2,\ \ c_2 = c_6 = 0,\ \ c_3 = c_5 = c_7 = -1.$

$$\textit{Solution}: (a)\ \mathbf{c} = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, \qquad \mathbf{f}^{(0)} = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \qquad \mathbf{f}^{(1)} = \begin{pmatrix} 2 \\ 0 \\ -2 \\ 0 \end{pmatrix}, \qquad \mathbf{f} = \mathbf{f}^{(2)} = \begin{pmatrix} 0 \\ 0 \\ 4 \\ 0 \end{pmatrix}.$$

10.14. ♡ Let $\mathcal{G}$ and $\widetilde{\mathcal{G}}$ be two weighted graphs with m and $\widetilde{m}$ nodes, and with weight matrices W and $\widetilde{W}$, respectively. The *Cartesian product graph* $\mathcal{G} \times \widetilde{\mathcal{G}}$, whose nodes are indexed by pairs (i,j) of nodes in each graph, so $1 \le i \le m$ and $1 \le j \le \widetilde{m}$, and whose weight matrix is denoted by $W \times \widetilde{W}$, with edge weights

$$(W \times \widetilde{W})_{(i,j),(k,\ell)} = \begin{cases} w_{ik}, & \text{if } j = \ell \\ \widetilde{w}_{j\ell}, & \text{if } i = k \\ 0, & \text{otherwise.} \end{cases}$$

Note that $W \times \widetilde{W}$ is an $m\widetilde{m} \times m\widetilde{m}$ matrix, and the nodal set for $\mathcal{G} \times \widetilde{\mathcal{G}}$ is the Cartesian product of the nodal sets of each graph. Let L and $\widetilde{L}$ denote the graph Laplacian matrices for $\mathcal{G}$ and $\widetilde{\mathcal{G}}$, respectively. For $\mathbf{u} \in \mathbb{R}^m$ and $\widetilde{\mathbf{u}} \in \mathbb{R}^{\widetilde{m}}$, let $\mathbf{u} \times \widetilde{\mathbf{u}} \in \mathcal{M}_{m \times \widetilde{m}}$ be the matrix with (i,j) entry $u_i \widetilde{u}_j$ which we identify in the usual manner with a vector in $\mathbb{R}^{m\widetilde{m}}$. Let $L \times \widetilde{L}$ denote the $m\widetilde{m} \times m\widetilde{m}$ graph Laplacian matrix on $\mathcal{G} \times \widetilde{\mathcal{G}}$.

(a) Show that if $\mathbf{u}$ is an eigenvector of L with eigenvalue λ, and $\widetilde{\mathbf{u}}$ is an eigenvector of $\widetilde{L}$ with eigenvalue $\widetilde{\lambda}$, then $\mathbf{u} \times \widetilde{\mathbf{u}}$ is an eigenvector of $L \times \widetilde{L}$ with eigenvalue $\lambda + \widetilde{\lambda}$.

(b) Do the eigenvectors obtained this way form a basis for $\mathbb{R}^{m\widetilde{m}}$?

Solution: (a) This is just a matter of unwrapping definitions. Note that the neighbors of the node (i,j) in $\mathcal{G} \times \widetilde{\mathcal{G}}$ are the nodes (k,ℓ) where either k is a neighbor of node i in $\mathcal{G}$ and $j = \ell$, or ℓ is a neighbor of node j in $\widetilde{\mathcal{G}}$ with $i = k$. Thus, the degree of node (i,j) in the Cartesian product graph is simply the sum of the degrees $d_i + \widetilde{d}_j$ of node i and node j in their respective graphs. Following this same logic,

$$(L \times \widetilde{L})(\mathbf{u} \times \widetilde{\mathbf{u}}) = (L\mathbf{u}) \times \widetilde{\mathbf{u}} + \mathbf{u} \times \widetilde{L}\widetilde{\mathbf{u}}.$$

The eigenvector condition yields

$$(L \times \widetilde{L})(\mathbf{u} \times \widetilde{\mathbf{u}}) = \lambda \mathbf{u} \times \widetilde{\mathbf{u}} + \widetilde{\lambda} \mathbf{u} \times \widetilde{\mathbf{u}} = (\lambda + \widetilde{\lambda})(\mathbf{u} \times \widetilde{\mathbf{u}}). \qquad \blacksquare$$

(b) The answer is yes. To see this, we just need to show that the $m\widetilde{m}$ eigenvectors of the form $\mathbf{u}_i \times \widetilde{\mathbf{u}}_j$, where $L\mathbf{u}_i = \lambda_i \mathbf{u}_i$ and $\widetilde{L}\,\widetilde{\mathbf{u}}_j = \widetilde{\lambda}_j \widetilde{\mathbf{u}}_j$, are linearly independent. To see this assume that

$$\sum_{i=1}^{m} \sum_{j=1}^{\widetilde{m}} c_{ij} (\mathbf{u}_i \times \widetilde{\mathbf{u}}_j) = \mathbf{0},$$

which can be arranged to read

$$\sum_{i=1}^{m} \mathbf{u}_i \times \mathbf{p}_i = \mathbf{0}, \qquad \text{where} \qquad \mathbf{p}_i = \sum_{j=1}^{\widetilde{m}} c_{ij} \widetilde{\mathbf{u}}_j.$$

In the more familiar matrix notation, where we identify a vector $\mathbf{v} \times \mathbf{w} \in \mathbb{R}^{m\widetilde{m}}$ with the $m \times \widetilde{m}$ matrix $\mathbf{v}\,\mathbf{w}^T$, this reads

$$\sum_{i=1}^{m} \mathbf{u}_i\,\mathbf{p}_i^T = \mathrm{O}.$$

Writing $\mathbf{p}_i = \left(p_{i,1}, \ldots, p_{i,\widetilde{m}}\right)^T$, the j-th column of the preceding equation is

$$\sum_{i=1}^{m} p_{i,j}\,\mathbf{u}_i = \mathbf{0}.$$

Since the $\mathbf{u}_i$ are linearly independent, we deduce that $p_{ij} = 0$ for all i, j, and so $\mathbf{p}_i = \mathbf{0}$ for all i. Using the definition of $\mathbf{p}_i$ and the linear independence of the $\widetilde{\mathbf{u}}_j$, we conclude that $c_{ij} = 0$ for all i, j, which establishes the claim. ■

10.16. ♡ Write a Python notebook to extend the signal denoising example with the DFT to two dimensional signals (i.e., images) and use it to denoise natural images corrupted by noise. The Python notebook from this section will be helpful to start out.

Solution:

Python Notebook: Solution to Exercise 10.16. (.ipynb)

Chapter 10

Neural Networks and Deep Learning

1.1. ♡ Explain why we can, without loss of generality, assume that all but the last activation function in a neural network can be taken to be non-affine.

Solution: If σ_k is an affine function, then, since a composition of affine functions is also an affine function, cf. (3.64), F_k would be an affine function of $\mathbf{z}_{k-1}$, and hence $\mathbf{p}_{k+1}$ would also be an affine function of $\mathbf{z}_{k-1}$. But we could then omit the $(k+1)$-st layer and generate the $(k+2)$-nd preactivation directly from the kth by an affine function.

1.2. ♡ Consider a *bias-free* ResNet, as in (10.15) but where all the bias vectors $\mathbf{b}_k$ are removed, i.e., we set $\mathbf{b}_k = 0$. Explain why any ResNet with input $\mathbf{x} \in \mathbb{R}^n$ has a bias-free version that gives the same outputs by taking $\begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} \in \mathbb{R}^{n+1}$ as input.

Solution: Set $\widetilde{\mathbf{x}} = \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix}$ and $\widetilde{W}_k = \begin{pmatrix} W_k & \mathbf{b}_k \\ \mathbf{0}^T & 1 \end{pmatrix}$ so that $\widetilde{W}_k \widetilde{\mathbf{x}} = \begin{pmatrix} W_k \mathbf{x} + \mathbf{b}_k \\ 1 \end{pmatrix}$. Thus, setting $\widetilde{V}_k = \begin{pmatrix} V_k & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix}$, and using (10.15), the bias-free ResNet layer

$$\widetilde{F}_k(\widetilde{\mathbf{x}}) = \widetilde{\mathbf{x}} + \widetilde{V}_k \sigma(\widetilde{W}_k \widetilde{\mathbf{x}}) = \begin{pmatrix} \mathbf{x} + V_k \sigma(W_k \mathbf{x} + \mathbf{b}_k) \\ 1 \end{pmatrix} = \begin{pmatrix} F_k(\mathbf{x}) \\ 1 \end{pmatrix}$$

coincides with the biased version (10.15) once the last entry of the vector is ignored.

1.3. ♡ Modify the Python notebooks from this section to use the loss (10.20) for classification instead of the softmax and negative log likelihood loss. Try different norms for the $\| F(\mathbf{x}_i) \|$ term. In particular, try the p norm with different values for p; in particular, try $p > 2$ (for example, $p = 5$). Are you able to get similar accuracy to the softmax composed with negative log likelihood loss?

Solution: We experimented with $2 \le p \le 5$. The accuracy is similar to softmax/cross entropy loss, though for large values, like $p = 5$, the results are actually very slightly better.

Python Notebook: Solution to Exercise 1.3. (.ipynb)

2.1.♡ Show that the Jacobian of the function $F(\mathbf{x}) = \sigma(W\mathbf{x} + \mathbf{b})$ equals $\mathbf{D}F(\mathbf{x}) = DW$, where $D = \operatorname{diag}\sigma'(W\mathbf{x} + \mathbf{b})$.

Solution: The i-th component of F is $\sigma\left(\displaystyle\sum_{k=1}^{n} w_{ik}x_k + b_i\right)$, and so its derivative with respect to x_j is $w_{ij}\,\sigma'\left(\displaystyle\sum_{k=1}^{n} w_{ik}x_k + b_i\right) = w_{ij}d_{ii}$, which is the (i,j) entry of DW. ■

2.4.♡ Prove Theorem 10.5. *Hint*: Mimic the proof of Theorem 10.3.

Solution: For $k = 0, \ldots, L$, let $G_k = \ell \circ F_L \circ \cdots \circ F_{k+1}$, so that $\ell\big(F(\mathbf{x})\big) = G_k(\mathbf{z}_k)$, where $\mathbf{z}_k = F_k(\mathbf{z}_{k-1}; \mathbf{w}_k)$ and $\mathbf{z}_0 = \mathbf{x}$. Since $G_k \circ F_k = G_{k-1}$, the chain rule yields

$$\begin{aligned}\mathbf{v}_{k-1} = \nabla_{\mathbf{z}_{k-1}}\ell\big(F(\mathbf{x})\big) &= \nabla_{\mathbf{z}_{k-1}}G_k\big(F_k(\mathbf{z}_{k-1}; \mathbf{w}_k)\big)\\ &= \mathbf{D}_{\mathbf{z}}F_k(\mathbf{z}_{k-1}, \mathbf{w}_k)^T\,\nabla_{\mathbf{z}_k}G_k = \mathbf{D}_{\mathbf{z}}F_k(\mathbf{z}_{k-1}; \mathbf{w}_k)^T\,\mathbf{v}_k,\end{aligned}$$

which establishes the first formula in (10.32). Similarly,

$$\nabla_{\mathbf{w}_k}\ell\big(F(\mathbf{x})\big) = \nabla_{\mathbf{w}_k}G_k\big(F_k(\mathbf{z}_{k-1}; \mathbf{w}_k)\big) = \mathbf{D}_{\mathbf{w}}F_k(\mathbf{z}_{k-1}, \mathbf{w}_k)^T\,\nabla_{\mathbf{z}_k}G_k = \mathbf{D}_{\mathbf{w}}F_k(\mathbf{z}_{k-1}; \mathbf{w}_k)^T\,\mathbf{v}_k,$$

proving the second formula. ■

2.5.♡ Implement a one hidden layer neural network in Python using only the `numpy` package. Use the formulas from Example 10.4 to compute the gradients of the network and implement plain gradient descent for training. Try your network on some of the basic classification examples from Section 10.1, including MNIST.

Solution: The notebook below implements a 1-hidden layer ReLU neural network with 64 hidden nodes in `numpy` for classification of MNIST digits using softmax and negative log likelihood loss. The code is written to look like PyTorch, though everything is explicitly written in `numpy`. The training is with vanilla stochastic gradient descent with a learning rate of $\alpha = 1$ and batch size of 128. Convergence is slower than with advanced optimizers like Adam, and the backpropagation runs slower as well. If you wait for about 50 epochs, the testing accuracy is around 95%, and continues to increase after that.

Python Notebook: Solution to Exercise 2.5. (.ipynb)

3.1. ♡ Modify the Python notebook from this section to use the Adam optimizer instead of Adadelta and see if you can obtain similar results. You may have to adjust the learning rate to ensure the loss decreases during training.

Solution: Change the optimizer line to read
```
optimizer = optim.Adam(model.parameters(), lr=learning_rate)
```
You will also have to change the learning rate to
```
learning_rate = 0.01
```
You will get an accuracy above 98%, but it may not be as high as the original notebook with the Adadelta optimizer.

3.2. ♡ How many parameters are in the convolutional part and the fully connected part of the CNN used for MNIST digit recognition in this section? Count just the weight matrices and ignore the biases, for simplicity.

Solution: The first convolutional layer has 32 3×3 filters applied to grayscale (single channel) images, which is $9 \times 32 = 288$ parameters. The second layer has 64 $3 \times 3 \times 32$ filters (since they are applied to 32 channel images from the first layer), which is $64 \times 9 \times 32 = 18432$ parameters. Thus, the convolutional layers have 18720 parameters. The fully connected network has two layers and 1024 hidden nodes, with 1600 inputs and 10 outputs. Thus, the number of parameters in the weight matrices is $1600 \times 1024 + 1024 \times 10 = 1648640$. Thus, there are about 88 times as many parameters in the fully connected part of the network, as compared to the convolutional part.

3.4. ♡ Modify the Python notebook from this this section so that the CNN can be trained for fully supervised classification on the CIFAR-10 data set, which has color images of size $32 \times 32 \times 3$, and train the network in a similar way as we did for MNIST (in particular, keep the size of the network roughly the same). What accuracy do you get?

Solution: The main thing we need to do is modify the architecture so the input layer allows three channels, and keep in mind that the images are 32×32, so after the two convolutional and max pooling layers, we'll have 6×6 images, not 5×5 as with MNIST, so we'll have $6 \times 6 \times 64 = 2304$ convolutional features. After 20 epochs of training, the accuracy will be around 74%, and it does not increase much upon training for longer. The best results on CIFAR-10 are above 99%, using either much larger CNN models, or transformers (see Section 10.5) [62]. The solution notebook is below.

Python Notebook: Solution to Exercise 3.4. (.ipynb)

4.1. ♡ What conditions could you place on $\mathbf{w}$ in Lemma 10.12 so that the result (10.42) remains true even when the eigenvalues of L_{sym} are not distinct?

Solution: If $\lambda_i = \lambda_j$ then you need $w_i = w_j$ in order define an interpolating polynomial (which now has smaller degree) satisfying $p(\lambda_i) = w_i$ for all i.

4.2. ♡ Show that the spectral graph convolution in Definition 10.11 can be written as

$$\mathbf{x} *_{\mathcal{G}} \mathbf{y} = \sum_{i=1}^{m} (\mathbf{q}_i \cdot \mathbf{x})\,(\mathbf{q}_i \cdot \mathbf{y})\,\mathbf{q}_i. \tag{10.48}$$

Solution: The i-th entry of $Q^T\mathbf{x}$ is $\mathbf{q}_i \cdot \mathbf{x}$ and similarly for $Q^T\mathbf{y}$. Thus the i-th entry of their Hadamard product is $(\mathbf{q}_i \cdot \mathbf{x})\,(\mathbf{q}_i \cdot \mathbf{y})$. The result thus follows from the column-based formula (3.21) for multiplying a matrix by a vector. ■

4.5. ♡ Implement a diffusion GCN, which uses diffusion (10.44) as the convolution in each layer. This has tunable parameters w_j that need to be learned. Write your code where the degree of the convolution k can be arbitrarily chosen. You may want to use `nn.Parameter` to define the weights w_j in the neural network. For simplicity, use the same convolution in each layer. Try this on the PubMed data set, and compare to the Python notebook from this section.

Solution: The Python notebook is below. You will generally find the result worse at low label rates for the more expressive diffusion convolutions.

Python Notebook: Solution to Exercise 4.5. (.ipynb)

5.1. ♡ Explain how the initial vector embedding of tokens into $\mathbb{R}^n$ can be viewed as a linear function, i.e., matrix multiplication, of the $M \times n$ token embedding matrix E with a matrix containing the one-hot encoding $\mathbf{e}_i$ of each token in the context window.

Solution: The token matrix E is the $M \times n$ matrix whose i-th row represents the vector embedding in $\mathbb{R}^n$ of the i-th token in the dictionary. Suppose we have a context with tokens indexed by $i_1, \ldots, i_m \in \{1, \ldots, M\}$. The corresponding tokens as row vectors are given by $\mathbf{e}_{i_1}^T E, \ldots, \mathbf{e}_{i_m}^T E$, which are simply the corresponding rows of E. Thus, the matrix X whose rows are the embeddings of the m tokens is the matrix $X = (\mathbf{e}_{i_1} \ \ldots \ \mathbf{e}_{i_m})^T E$.

5.3. ♡ Let π be a permutation of the integers $1, \ldots, m$. (a) Show that if $\pi(i) = i$, then the output $\mathbf{z}_i$ of the attention layer (10.52) is unchanged by permuting the tokens with π, that is

$$\mathbf{z}_i = \frac{1}{d_{\pi(i)}} \sum_{j=1}^{m} a_{\pi(i)\,\pi(j)}\, V\,\mathbf{x}_{\pi(j)}.$$

(b) Show that the same is true for masked self-attention described in Remark 10.19 provided $\pi(i) = i$ and $\pi(j) \le i$ for all $j \le i$.

Solution: (a) If $\pi(i) = i$ then

$$\mathbf{z}_i = \frac{1}{d_i} \sum_{j=1}^{m} a_{ij}\, V\, \mathbf{x}_j = \frac{1}{d_i} \sum_{j=1}^{m} a_{i\,\pi(j)}\, V\, \mathbf{x}_{\pi(j)},$$

where we simply re-indexed the sum in the second inequality.

(b) The condition on π implies that π restricted to $\{1, \ldots, i-1\}$ is a permutation of the first $i-1$ tokens. Thus, we can apply the same argument as in (a) to find that masked self attention is

$$\mathbf{z}_i = \frac{1}{d_i} \sum_{j \le i} a_{ij}\, V\, \mathbf{x}_j = \frac{1}{d_i} \sum_{j \le i} a_{i\,\pi(j)}\, V\, \mathbf{x}_{\pi(j)}.$$

6.1. ♡ Prove that the space of piecewise affine functions $\mathcal{A}$ forms a vector space. Explain in detail why it is also affine invariant.

Solution: Any constant multiple of an affine function is also affine: $c\,(a\,x+b) = (c\,a)\,x+(c\,b)$. Consequently, a constant multiple of a piecewise affine function is also piecewise affine with the same nodes. Similarly, the sum of two affine functions is affine: $(a\,x + b) + (c\,x + d) = (a + c)\,x + (b + d)$. Thus, the sum of two piecewise affine functions that have the same nodes is also piecewise affine. If the two piecewise affine functions have different nodes, then the nodes of their sum will be the union of their nodes. Further, they are both piecewise affine on the union of nodes (adding one of more new nodes does not alter the status of the function being piecewise affine) and hence, by the preceding observation, the sum is piecewise affine on the union of the nodes.

To prove affine invariance, first if $x = c\,y+d$ with $c \neq 0$, then $a\,x+b = a\,c\,y+(a\,d+b)$ is also affine. We conclude that applying such an affine transformation to a piecewise affine function with nodes x_i produces a piecewise affine function with nodes y_i satisfying $x_i = c\,y_i + d$.

6.4. ♡ Prove that if $u \in \mathrm{Lip}(\mathbb{R})$, there exists a 2π-periodic function $\widetilde{u} \in \mathrm{Lip}(\mathbb{R})$ such that $u(x) = \widetilde{u}(x)$ for all $0 \le x \le 1$.

Solution: Let $a = u(0^+)$ and $b = u(1^-)$. Set $\widetilde{u}$ be the 2π periodic extension of the function

$$\widetilde{u}(x) = \begin{cases} u(x), & 0 \le x \le 1, \\ b + (a-b)\,(x-1)/(2\pi - 1), & 1 \le x \le 2\pi, \end{cases} \qquad \text{where} \qquad \mu = \frac{a-b}{2\pi - 1},$$

which is defined so that $\widetilde{u}$ is continuous on all of $\mathbb{R}$. If $0 \le y \le x \le 1$, then

$$|\, \widetilde{u}(x) - \widetilde{u}(y)\,| = |\, u(x) - u(y)\,| \le \lambda\,|\, x - y\,|,$$

where λ is the Lipschitz constant of u. If $0 \le y \le 1 \le x \le 2\pi$, then

$$|\, \widetilde{u}(x) - \widetilde{u}(y)\,| \le |\, \widetilde{u}(x) - \widetilde{u}(1)\,| + |\, \widetilde{u}(1) - \widetilde{u}(y)\,| \le \mu\,(x-1) + \lambda\,(1-y) \le \nu\,|\, x - y\,|,$$

where $\nu = \max\{\,\lambda, |\,\mu\,|\,\}$. A similar proof works if $1 - 2\pi \le y \le 0 \le x \le 1$. If $1 \le y \le x \le 2\pi$, then $|\, \widetilde{u}(x) - \widetilde{u}(y)\,| = |\,\mu\,|\,|\, x - y\,|$. Finally, for any $\widetilde{x}, \widetilde{y} \in \mathbb{R}$, we can find $j, k \in \mathbb{Z}$ such that $x = \widetilde{x} - 2\,j\,\pi$ and $y = \widetilde{y} - 2\,k\,\pi$ satisfy one of the preceding inequalities and, moreover, $|\, x - y\,| \le |\, \widetilde{x} - \widetilde{y}\,|$. Then

$$|\, \widetilde{u}(\widetilde{x}) - \widetilde{u}(\widetilde{y})\,| = |\, \widetilde{u}(x) - \widetilde{u}(y)\,| \le \nu\,|\, x - y\,| \le \nu\,|\, \widetilde{x} - \widetilde{y}\,|,$$

completing the proof. ■

6.5. ♡ Prove that, for each $k \geq 1$, the function (10.85) is continuous, piecewise affine, and satisfies (10.86).

Solution: The composition of two affine functions is affine; see (3.64). Therefore, the composition of two continuous piecewise affine functions is continuous piecewise affine. Finally, we prove (10.86) by induction on k, the case $k = 1$ being obvious from the construction of the hat function $g(x)$. First note that

$$g_k\left(\frac{j}{2^{k+1}}\right) = \begin{cases} \frac{1}{2}, & j \text{ odd}, \\ 0, & j \equiv 0 \bmod 4, \\ 1, & j \equiv 2 \bmod 4, \end{cases}$$

where the first case follows since the value of an affine function at the midpoint of an interval is the average of its values at the ends. Thus, by the definition of the hat function,

$$g_{k+1}\left(\frac{j}{2^{k+1}}\right) = g\left(g_k\left(\frac{j}{2^{k+1}}\right)\right) = \begin{cases} g\left(\frac{1}{2}\right), & j \text{ odd}, \\ g(0), & j \equiv 0 \bmod 4, \\ g(1), & j \equiv 2 \bmod 4, \end{cases} = \begin{cases} 1, & j \text{ odd}, \\ 0, & j \text{ even}. \end{cases}$$ ■

6.6. ♡ (a) Use polar coordinates to prove that, for any $a > 0$,

$$\iint_{\mathbb{R}^2} e^{-a(x^2+y^2)}\, dx\, dy = \frac{\pi}{a}. \tag{10.99}$$

(b) Explain why

$$\int_{-\infty}^{\infty} e^{-a x^2}\, dx = \sqrt{\frac{\pi}{a}}. \tag{10.100}$$

Solution: (a) Setting $x = r\cos\theta$, $y = r\sin\theta$, we have $dx\, dy = r\, dr\, d\theta$, and hence the double integral is

$$\begin{aligned} \iint_{\mathbb{R}^2} e^{-a(x^2+y^2)}\, dx\, dy &= \int_{-\pi}^{\pi} \int_0^{\infty} r\, e^{-a r^2}\, dr\, d\theta \\ &= 2\pi \int_0^{\infty} r\, e^{-a r^2}\, dr = -\frac{\pi}{a}\, e^{-a r^2}\Big|_{r=0}^{\infty} = \frac{\pi}{a}. \end{aligned}$$ ■

(b) By part (a),

$$\begin{aligned} \frac{\pi}{a} &= \iint_{\mathbb{R}^2} e^{-a(x^2+y^2)}\, dx\, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-a x^2} e^{-a y^2}\, dy\, dx \\ &= \left(\int_{-\infty}^{\infty} e^{-a x^2}\, dx\right)\left(\int_{-\infty}^{\infty} e^{-a y^2}\, dy\right) = \left(\int_{-\infty}^{\infty} e^{-a x^2}\, dx\right)^2. \end{aligned}$$

Taking square roots of both sides proves the identity. ■

6.8. ♡ Let $u \in \mathrm{Lip}(\mathbb{R})$, and let g be a piecewise affine interpolant of u on an interval $[a, b]$. Prove that $\mathrm{Lip}_{[a,b]}(g) \le \mathrm{Lip}_{[a,b]}(u)$.

Solution: Suppose $a_i < a_{i+1}$ are two consecutive nodes of g, so $g(a_i) = u(a_i)$ and $g(a_{i+1}) = u(a_{i+1})$, and hence

$$g(x) = u(a_i) + \frac{u(a_{i+1}) - u(a_i)}{a_{i+1}) - a_i}\,(x - a_i) \qquad \text{for} \qquad a_i \le x \le a_{i+1}.$$

Applying the Lipschitz estimate on u, we bound

$$|\,g'(x)\,| = \frac{|\,u(a_{i+1}) - u(a_i)\,|}{a_{i+1}) - a_i} \le \mathrm{Lip}_{[a,b]}(u) \qquad \text{for} \qquad a_i \le x \le a_{i+1}.$$

Since this holds on all subintervals of $[a, b]$, and a piecewise affine function is piecewise continuously differentiable, Proposition 6.59 implies $\mathrm{Lip}_{[a,b]}(g) \le \mathrm{Lip}_{[a,b]}(u)$. ■

6.9. ♡ Suppose $u\colon [a, b] \to [c, d]$. Let g be a piecewise affine interpolant of u on $[a, b]$. Prove that $g\colon [a, b] \to [c, d]$.

Solution: On each interval where g is affine, it is bounded from above and below by its values at the endpoints, where it has the same values as u. Thus the minimum and maximum values of g on $[a, b]$ are bounded by the minimum and maximum values of u, which are bounded from below and above by c and d, respectively.

6.10. ♡ Write down the hat functions corresponding to unequally spaced nodes, and prove that they still form a partition of unity.

Solution: Let $a_0, \ldots, a_n$ denote the nodes. Then $\theta_i(x)$ is the piecewise affine function interpolating $\theta_i(a_i) = 1$ and $\theta_i(a_j) = 0$ for $j \ne i$. Thus, on the interval $a_i \le x \le a_{i+1}$, we have

$$\theta_i(x) = \frac{x - a_{i+1}}{a_i - a_{i+1}}, \qquad \theta_{i+1}(x) = \frac{x - a_i}{a_{i+1} - a_i}, \qquad \theta_j(x) = 0, \qquad j \ne i, i+1,$$

and hence, for such x,

$$\sum_{j=0}^{n} \theta_j(x) = \theta_i(x) + \theta_{i+1}(x) = \frac{x - a_{i+1}}{a_i - a_{i+1}} + \frac{x - a_i}{a_{i+1} - a_i} = 1.$$ ■

Chapter 11

Advanced Optimization

1.1. ♡ Prove the inequality (11.6).

Solution: First note that for $x \in [0,1]$, we have $F(x) = 0 = F'(x)$, so (11.6) clearly holds. For $x < 0$, we have $F(x) = \frac{1}{2}x^2$ and $F'(x)^2 = x^2$, so (11.6) is satisfied there. For $x > 1$, we have $F(x) = \frac{1}{2}(x-1)^2$ and $F'(x) = x - 1$, therefore (11.6) holds again. ■

1.3. ♡ Prove Theorem 11.2.
Hint: The proof is nearly identical to Theorem 6.68. Use Lemma 6.64 and Exercise 1.2.

Solution: Using the inequality (6.129) and then the PL inequality (11.2), we find

$$F(\mathbf{x}_{k+1}) - F^\star \le F(\mathbf{x}_k) - F^\star - \frac{\alpha}{2}\,\|\nabla F(\mathbf{x}_k)\|^2 \le (1-\alpha\mu)\left[\,F(\mathbf{x}_k) - F^\star\,\right]$$

By Exercise 1.2 and the restriction $\alpha \le \mathrm{Lip}(\nabla F)^{-1}$ we have $1 - \alpha\mu \ge 0$. This allows us to iterate this final inequality to produce (11.3). ■

1.4. ♡ Show that the function F defined in (11.5) is not strongly convex.

Solution: Set $G(x) = F(x) - \mu x^2$. If $0 < x < 1$ and $0 < t < 1$, then $0 < (1-t)x < 1$ and $(1-t)^2 < (1-t)$. Thus,

$$G\big(\,t\cdot 0 + (1-t)\,x\,\big) = G\big(\,(1-t)\,x\,\big) = -\mu(1-t)^2x^2 > -\mu(1-t)\,x^2 = t\,G(0) + (1-t)\,G(x),$$

thus violating convexity. Alternatively, we can check that $F'(x) = 0$ for $0 < x < 1$ so F is not strictly nor strongly convex. ■

2.1. ♡ Suppose $\beta > 1$ in (11.15). For which initial conditions x_0, x_1 do the iterates *not* become unbounded as $k \to \infty$?

Solution: When $x_0 = x_1$ since then $x_k = x_0$ for all k. Indeed, the solution remains bounded if and only if the initial vector $\mathbf{c}_1 = x_0\,(\,1,1\,)^T$ is a multiple of the eigenvector $(\,1,1\,)^T$ for the eigenvalue $\lambda_1 = 1$.

2.3. ♡ Suppose that H is only positive semidefinite in Theorem 11.10. Assume $\mathbf{b} \in \operatorname{img} H$ so that there exists $\mathbf{x}^\star$ such that $H\mathbf{x}^\star = \mathbf{b}$. Let UU^T be the orthogonal projection matrix onto $\operatorname{img} H$. Show that for any choice of $\mathbf{x}^\star$ we have

$$\| UU^T(\mathbf{x}_k - \mathbf{x}^\star) \| \le \sqrt{2}\,(\sqrt{\beta} + \varepsilon)^k \, \| UU^T(\mathbf{x}_0 - \mathbf{x}^\star) \| \tag{11.21}$$

using all the same conditions as Theorem 11.10, except that $\lambda_{min}(H)$ is replaced by the smallest positive eigenvalue in (11.17). Can you prove anything about the convergence of $\mathbf{x}_k$ to $\mathbf{x}^\star$?

Solution: Let U be the $n \times k$ matrix whose columns are the orthonormal eigenvectors of H corresponding to non-zero eigenvalues. Then UU^T is the projection matrix onto $\operatorname{img} H$. The $k \times k$ matrix $U^T HU$ has the same non-zero eigenvalues as H, except the eigenvectors are $\mathbf{e}_1, \ldots, \mathbf{e}_k$; indeed $U^T HU\mathbf{e}_i = \lambda_i U^T U\mathbf{e}_i = \lambda_i \mathbf{e}_i$, since $U\mathbf{e}_i$ is the i-th column of U, which is an eigenvalue of H whose eigenvalue we denote by λ_i. Thus, $U^T HU$ is positive definite symmetric, and $\lambda_{min}(U^T HU)$ is the smallest nonzero eigenvalue of H, while $\lambda_{max}(U^T HU) = \lambda_{max}(H)$. We can take the heavy ball method for H, multiply by U^T on both sides, and use $H = HUU^T$ to obtain

$$U^T\mathbf{x}_{k+1} = U^T\mathbf{x}_k - \alpha(U^T HUU^T\mathbf{x}_k - U^T\mathbf{b}) + \beta(U^T\mathbf{x}_k - U^T\mathbf{x}_{k-1}).$$

Writing $S = U^T HU$ and $\mathbf{z}_k = U^T\mathbf{x}_k$ we have

$$\mathbf{z}_{k+1} = \mathbf{z}_k - \alpha(S\mathbf{x}_k - U^T\mathbf{b}) + \beta(\mathbf{z}_k - \mathbf{z}_{k-1}).$$

This is the heavy ball method for S and $U^T\mathbf{b}$, so we can apply Theorem 11.10 to obtain

$$\| \mathbf{z}_k - U^T\mathbf{x}^\star \| \le \sqrt{2}\,(\sqrt{\beta} + \varepsilon)^k \, \| \mathbf{z}_0 - U^T\mathbf{x}^\star \|,$$

where in the condition (11.18) we note that the argument above shows that $\lambda_{min}(S)$ is the smallest nonzero eigenvalue of H, while $\lambda_{max}(S) = \lambda_{max}(H)$. We now use the fact that $\|UU^T\mathbf{x}\| = \|U^T\mathbf{x}\|$ (square both sides and expand using $U^T U = \mathrm{I}$) to obtain

$$\| UU^T(\mathbf{x}_k - \mathbf{x}^\star) \| \le \sqrt{2}\,(\sqrt{\beta} + \varepsilon)^k \, \| UU^T(\mathbf{x}_0 - \mathbf{x}^\star) \|.$$

We cannot prove anything about the convergence of $\mathbf{x}_k \to \mathbf{x}^\star$ without selecting a particular $\mathbf{x}^\star$, since it is not unique. We can choose $\mathbf{x}^\star$ to be the orthogonal projection of $\mathbf{x}_0$ onto the solution space $\{\mathbf{x} \,|\, H\mathbf{x} = \mathbf{b}\}$, which satisfies $\mathbf{x}_0 - \mathbf{x}^\star \in (\ker H)^\perp = \operatorname{img} H$. It follows that $\mathbf{x}_k - \mathbf{x}^\star \in \operatorname{img} H$ for all k, and so we can remove the projection matrix UU^T and obtain

$$\| \mathbf{x}_k - \mathbf{x}^\star \| \le \sqrt{2}\,(\sqrt{\beta} + \varepsilon)^k \, \| \mathbf{x}_0 - \mathbf{x}^\star \|,$$

and thus $\mathbf{x}_k \to \mathbf{x}_0$. ∎

3.1. ♡ Prove by induction that if the Krylov subspaces satsify $V_{j+1} = V_j$ for some $j \geq 0$, then $V_k = V_j$ for all $k \geq j$.

Solution: By assumption, $A^j \mathbf{v} \in V_j$, and thus can be written as a linear combination

$$A^j \mathbf{v} = c_1 \mathbf{v} + c_2 A \mathbf{v} + \ \cdots \ + c_{j-1} A^{j-2} \mathbf{v} + c_j A^{j-1} \mathbf{v} \in V_j$$

for some scalars $c_1, \ldots, c_j$. Thus,

$$\begin{aligned} A^{j+1} \mathbf{v} &= c_1 A \mathbf{v} + c_2 A^2 \mathbf{v} + \ \cdots \ + c_{j-1} A^{j-1} \mathbf{v} + c_j A^j \mathbf{v} \\ &= c_j c_1 \mathbf{v} + (c_1 + c_j c_2) A \mathbf{v} + \cdots + (c_{j-2} + c_j c_{j-1}) A^{j-2} \mathbf{v} + (c_{j-1} + c_j^2) A^{j-1} \mathbf{v} \in V_j \end{aligned}$$

also, which implies $V_{j+2} = V_j$. The general induction step follows. ■

3.2. ♡ Let $T_k(x)$ denote the k-th order Chebyshev polynomial (11.28).

(a) Show that both formulas in (11.28) satisfy the recursive formula (10.46). *Hint*: Use the trigonometric identity $\cos(x+y) + \cos(x-y) = 2 \cos x \cos y$.

(b) Explain why both formulas define the same polynomial of degree k.

(c) Write down expressions for the Chebyshev polynomials $T_3(x)$ and $T_4(x)$.

(d) For $\kappa > 1$, show that $\quad T_k \left(\dfrac{\kappa+1}{\kappa-1} \right) = \dfrac{1}{2} \left[\left(\dfrac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1} \right)^k + \left(\dfrac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} \right)^k \right]$.

Solution: (a) By the indicated trigonometric identity

$$\cos(k\, y) = 2 \cos y \, \cos\big((k-1)\, y \big) - \cos\big((k-2)\, y \big).$$

Setting $y = \arccos x$, this becomes the recursive formula (10.46):

$$\cos(k \arccos x\,) = 2\, x \cos\big((k-1) \arccos x\,\big) - \cos\big((k-2) \arccos x\,\big).$$

Similarly,

$$\begin{aligned} 2\, x \left(x \pm \sqrt{x^2-1}\, \right)^{k-1} - \left(x \pm \sqrt{x^2-1}\, \right)^{k-2} &= \left(2\, x^2 \pm 2\, x \sqrt{x^2-1} - 1 \right) \left(x \pm \sqrt{x^2-1}\, \right)^{k-2} \\ &= 2 \left(x \pm \sqrt{x^2-1}\, \right)^k. \end{aligned}$$

Summing the $+$ and the $-$ equations and dividing by 2 establishes the recursive formula (10.46) in this case.

(b) Both formulas give $T_0(x) = 1$ and $T_1(x) = x$, and hence, in view of (10.46), define the same function $T_k(x)$ for all k. Moreover, by induction, $T_{k-1}(x)$ is a polynomial of degree $k-1$, which implies $2\, x\, T_{k-1}(x)$ is a polynomial of degree k, while $T_{k-2}(x)$ is a polynomial of degree $k-2$, and hence the sum $T_k(x) = 2\, x\, T_{k-1}(x) - T_{k-2}(x)$ is a polynomial of degree k.

(c) $T_3(x) = 4\, x^3 - 3\, x, \quad T_4(x) = 8\, x^4 - 8\, x^2 + 1$.

(d) The result is easily seen to be valid for $k = 0, 1$. We prove the general case by induction using the recursive formula (10.46) with $x = (\kappa+1)/(\kappa-1)$. Assuming the validity of the formula for $k-1$ and $k-2$, the right hand side of (10.46), namely,

$$2 \left(\frac{\kappa+1}{\kappa-1} \right) T_{k-1} \left(\frac{\kappa+1}{\kappa-1} \right) - T_{k-2} \left(\frac{\kappa+1}{\kappa-1} \right)$$

is the sum of the following two terms:

$$\begin{aligned} \left[\left(\frac{\kappa+1}{\kappa-1} \right) \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1} \right) - \frac{1}{2} \right] \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1} \right)^{k-2} &= \frac{1}{2} \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1} \right)^k, \\ \left[\left(\frac{\kappa+1}{\kappa-1} \right) \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} \right) - \frac{1}{2} \right] \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} \right)^{k-2} &= \frac{1}{2} \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} \right)^k. \end{aligned}$$

Adding them produces $T_k\big((\kappa+1)/(\kappa-1) \big)$, and thus completes the induction step. ■

3.5. ♡ Choose a polynomial $q(\lambda)$ in the proof of Theorem 11.15 to show that if H has at exactly k distinct eigenvalues then $\mathbf{x}_k = \mathbf{x}^\star$.

Solution: Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues. Set $q(\lambda) = \dfrac{(\lambda - \lambda_1)\cdots(\lambda - \lambda_k)}{(-1)^k \lambda_1 \cdots \lambda_k}$, so that $q(\lambda)$ has degree k, and $q(\lambda_i) = 0$ for all $i = 1, \ldots, k$, while $q(0) = 1$. Thus $q(\lambda)$ satisfies the conditions for (11.27) to be valid. However, the left hand side of the inequality is 0, and hence $F(\mathbf{x}_k) = F(\mathbf{x}^\star)$, which, by uniqueness of the minimizer, implies $\mathbf{x}_k = \mathbf{x}^\star$.

4.2. ♡ Consider Nesterov's accelerated gradient descent (11.33) where the gradient ∇F is defined with respect to an inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^n$. Formulate and prove a version of Theorem 11.18 in this setting.

Solution: There are two ways to do this problem. One is to go through the proof of Theorem 11.18 and replace all dot products by inner products, and Euclidean norms by the induced inner product norm. Everything goes through the same way, except the Lipschitz constant of ∇F is interpreted in the induced norm and the norm on the right hand side in the rate also uses the induced norm.

Alternatively, we can make a change of variables in Theorem 11.18, which we do here. Nesterov's accelerated gradient descent in the inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T C \mathbf{y}$ is

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \beta_k(\mathbf{x}_k - \mathbf{x}_{k-1}) - \alpha C^{-1} \nabla F\big(\mathbf{x}_k + \beta_k(\mathbf{x}_k - \mathbf{x}_{k-1})\big).$$

Let $G(\mathbf{x}) = F(A\mathbf{x})$, where A is a symmetric and nonsingular matrix to be determined. Then $\nabla G(\mathbf{x}) = A \nabla F(A\mathbf{x})$ and so

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \beta_k(\mathbf{x}_k - \mathbf{x}_{k-1}) - \alpha C^{-1} A^{-1} \nabla G\big(A^{-1}(\mathbf{x}_k + \beta_k(\mathbf{x}_k - \mathbf{x}_{k-1}))\big).$$

Setting $\mathbf{z}_k = A^{-1}\mathbf{x}_k$ and multiplying by A^{-1} on both sides yields

$$\mathbf{z}_{k+1} = \mathbf{z}_k + \beta_k(\mathbf{z}_k - \mathbf{z}_{k-1}) - \alpha A^{-1} C^{-1} A^{-1} \nabla G\big(\mathbf{z}_k + \beta_k(\mathbf{z}_k - \mathbf{z}_{k-1})\big).$$

Now we see a good choice is $A = C^{-1/2}$, in which case $G(\mathbf{x}) = F(C^{-1/2}\mathbf{x})$, $\mathbf{z}_k = C^{1/2}\mathbf{x}_k$, and

$$\mathbf{z}_{k+1} = \mathbf{z}_k + \beta_k(\mathbf{z}_k - \mathbf{z}_{k-1}) - \alpha \nabla G\big(\mathbf{z}_k + \beta_k(\mathbf{z}_k - \mathbf{z}_{k-1})\big)$$

is the usual Nesterov method in the Euclidean dot product. Thus, by Theorem 11.18,

$$F(\mathbf{x}_k) - F(\mathbf{x}^\star) \le \frac{2\,\|\mathbf{x}_0 - \mathbf{x}^\star\|_C^2}{\alpha(k-1)^2} \qquad \text{for} \qquad 0 < \alpha \le \mathrm{Lip}(\nabla G)^{-1}.$$

4.3. ♡ Implement Nesterov's accelerated gradient descent on a convex, but not strongly convex, function F in `numpy` in Python. Compare it with ordinary gradient descent and with the heavy ball method.

Solution:

Python Notebook: Solution to Exercise 4.3. (.ipynb)

5.1. ♡ Show that (11.54) implies (11.57).

Solution: We compute

$$\begin{aligned}\mathbb{E}_k\left(\xi_k\cdot\nabla F_{i_k}(\mathbf{x}_k)\right) &= \mathbb{E}_k\left(\|\nabla F_{i_k}(\mathbf{x}_k)\|^2-\nabla F(\mathbf{x}_k)\cdot\nabla F_{i_k}(\mathbf{x}_k)\right)\\ &= \mathbb{E}_k\|\nabla F_{i_k}(\mathbf{x}_k)\|^2-\nabla F(\mathbf{x}_k)\cdot\mathbb{E}_k\nabla F_{i_k}(\mathbf{x}_k)\le\|\nabla F(\mathbf{x}_k)\|^2+\sigma^2-\|\nabla F(\mathbf{x}_k)\|^2=\sigma^2.\end{aligned}$$

5.4. ♡ Let $\alpha_k=\alpha/(k+1)^p$ for $0<p<\frac{1}{2}$. Show that

$$\sum_{j=0}^{k-1}\alpha_j^2\le\frac{\alpha^2\,(k^{1-2p}-2p)}{1-2p}.$$

What happens when $p=\frac{1}{2}$?

Solution: We compute

$$\sum_{j=0}^{k-1}\alpha_j^2=\alpha^2\sum_{j=0}^{k-1}\frac{1}{(j+1)^{2p}}\le\alpha^2+\alpha^2\int_0^{k-1}\frac{1}{(x+1)^{2p}}\,dx\le\alpha^2+\frac{\alpha^2}{1-2p}(k^{1-2p}-1).$$

When $p=\frac{1}{2}$,

$$\sum_{j=0}^{k-1}\alpha_j^2\le\alpha^2+\alpha^2\int_0^{k-1}\frac{1}{x+1}\,dx\le\alpha^2\,(\log k+1)\,.$$ ■

5.6. ♡ Let $\alpha_k=\alpha/(k+1)^p$ where $0<p\le 1$. Use the previous three exercises and Theorem 11.22 to establish convergence rates for SGD with this choice of time step (use big O notation for simplicity). For which values of p is the method convergent? What happens when $p>1$?

Solution: By Theorem 11.22, the convergence rate is the larger of the two quantities $\mathrm{O}(1/(k\,\overline{\alpha}_k))$ and $\mathrm{O}(\varepsilon_k)$. When $0<p<\frac{1}{2}$ the rate is $\mathrm{O}(k^{-p})$. When $p=\frac{1}{2}$ the rate is $\mathrm{O}(\log k/\sqrt{k})$ as was worked out in this section. When $\frac{1}{2}<p<1$ the rate is $\mathrm{O}(k^{1-p})$. When $p=1$ the rate is $\mathrm{O}(1/\log k)$. In all cases $0<p\le 1$, SGD is convergent, with the fastest rate occurring when $p=\frac{1}{2}$. When $p>1$, the quantity $k\,\overline{\alpha}_k$ is bounded and does not grow to infinity, so SGD is not convergent. ■

6.1. ♡ (*a*) Solve the initial value problem

$$c''(t) + 2\,b\,c'(t) + \lambda\,c(t) = 0, \qquad c(0) = c_0, \qquad c'(0) = 0,$$

where $b > 0$, $\lambda > 0$, and $c_0 \in \mathbb{R}$.
(*b*) Show that $|\,c(t)\,| \le |\,c_0\,|\,(1 + b\,t)\,e^{-b\,t}$ whenever $b^2 \le \lambda$. *Hint*: Use $\sin x \le x$.
(*c*) What can you say about the solution if $b^2 > \lambda$?

Solution: (a) The solution is

$$c(t) = \begin{cases} c_0\,e^{-b\,t}\left[\cos\!\left(t\sqrt{\lambda - b^2}\,\right) + \dfrac{b}{\sqrt{\lambda - b^2}}\sin\!\left(t\sqrt{\lambda - b^2}\,\right)\right], & b^2 < \lambda, \\ c_0\,(1 + b\,t)\,e^{-b\,t} & b^2 = \lambda, \\ \dfrac{c_0}{2}\left[\left(1 + \dfrac{b}{\sqrt{b^2 - \lambda}}\right)e^{-(b - \sqrt{b^2 - \lambda})\,t} + \left(1 - \dfrac{b}{\sqrt{b^2 - \lambda}}\right)e^{-(b + \sqrt{b^2 - \lambda})\,t}\right], & b^2 > \lambda. \end{cases}$$

(*b*) When $b^2 = \lambda$, the inequality is an equality. When $b^2 < \lambda$, bounding $\cos x \le 1$ and $\sin x \le x$ in the solution formula in part (a) proves the bound.

(*c*) When $b^2 > \lambda$, we have $|\,c(t)\,| \le |\,c_0\,|\,e^{-(b - \sqrt{b^2 - \lambda})\,t}$, so the decay rate is slower.

6.5. ♡ Define $\beta_k = 1 - \sqrt{\alpha}\,F\!\left(k\sqrt{\alpha}\,\right)$. Prove that β_k is independent of α if and only if $F(t) = r/t$ for some $r \in \mathbb{R}$. *Hint*: Differentiate the expression for β_k with respect to α.

Solution: If β_k does not depend on α, then

$$0 = \frac{d\beta_k}{d\alpha} = -\,\frac{F\!\left(k\sqrt{\alpha}\,\right)}{2\sqrt{\alpha}} - \frac{k\,F'\!\left(k\sqrt{\alpha}\,\right)}{2} = -\,\frac{k\sqrt{\alpha}\,F'\!\left(k\sqrt{\alpha}\,\right) + F\!\left(k\sqrt{\alpha}\,\right)}{2\sqrt{\alpha}}.$$

Setting $t = k\sqrt{\alpha}$, we find that the function $F(t)$ must satisfy the ordinary differential equation

$$0 = t\,F'(t) + F(t) = \frac{d}{dt}\big[\,t\,F(t)\,\big], \qquad \text{and hence} \qquad t\,F(t) = r \qquad \text{is a constant.}$$ ■

6.6. ♡ Let $\mathbf{x}(t)$ be twice continuously differentiable and satisfy (11.90). Assume that ∇F is Lipschitz continuous. Show that $\mathbf{x}'(0) = \mathbf{0}$ and $\mathbf{x}''(0) = -\frac{1}{4}\,\nabla F(\mathbf{x}(0))$.
Hint: For $\mathbf{x}''$, use (11.91).

Solution: By (11.90) we have, for $0 \le t \le 1$,

$$\|\mathbf{x}'(t)\| \le \frac{1}{t^3}\int_0^t s^3\|\nabla F(\mathbf{x}(s))\|\,ds \le \frac{M_t}{4}t, \qquad \text{where} \qquad M = \max_{0 \le t \le 1}\|\nabla F(\mathbf{x}(t))\|.$$

Therefore $\mathbf{x}'(0) = \lim_{t \to 0}\mathbf{x}'(t) = \mathbf{0}$. Keeping track of the error terms in (11.91),

$$\mathbf{x}'(t) = -\frac{t}{4}\,\nabla F(\mathbf{x}(t)) + \mathrm{O}(t^2), \qquad \text{and so} \qquad \frac{3}{t}\,\mathbf{x}'(t) = -\frac{3}{4}\,\nabla F(\mathbf{x}(t)) + \mathrm{O}(t),$$

for t near zero. Plugging this into (11.89) yields

$$\mathbf{x}''(t) - \tfrac{3}{4}\,\nabla F(\mathbf{x}(t)) + \mathrm{O}(t) = -\,\nabla F(\mathbf{x}(t)).$$

Thus, sending $t \to 0$ produces $\mathbf{x}''(0) = -\,\frac{1}{4}\,\nabla F(\mathbf{x}(t))$. ■

7.1. ♡ Given $\mathrm{O} \neq X \in \mathcal{M}_{m\times n}$, show that $\| X^T X\mathbf{u} \| \geq \sigma_{min}(X) \| X\mathbf{u} \|$ for all $\mathbf{u} \in \mathbb{R}^n$.

Solution:

Let $X = \sum_{i=1}^{r} \sigma_i \mathbf{p}_i \mathbf{q}_i^T$ be its singular value decomposition, so that $X^T = \sum_{i=1}^{r} \sigma_i \mathbf{q}_i \mathbf{p}_i^T$ is the singular value decomposition of X^T. Since $X\mathbf{u} \in \operatorname{img} X$ we can write

$$X\mathbf{u} = \sum_{j=1}^{r} c_j \mathbf{p}_j, \qquad \text{and hence} \qquad X^T X = \left(\sum_{i=1}^{r} \sigma_i \mathbf{q}_i \mathbf{p}_i\right)^T \left(\sum_{j=1}^{r} c_j \mathbf{p}_j\right) = \sum_{i=1}^{r} \sigma_i c_i \mathbf{q}_i,$$

where we used orthonormality of the $\mathbf{p}_i$ in the last equality. Since the $\mathbf{q}_i$ are also orthonormal,

$$\| X^T X\mathbf{u} \|^2 = \sum_{i=1}^{r} \sigma_i^2 c_i^2 \geq \sigma_r^2 \sum_{i=1}^{r} c_i^2 = \sigma_{min}(X)^2 \| X\mathbf{u} \|^2.$$ ■

7.2. ♡ Suppose F has the form

$$F(\mathbf{x};\mathbf{w}) = \sum_{k=1}^{m} w_k \Phi_k(\mathbf{x}), \tag{11.110}$$

where each $\Phi_k : \mathbb{R}^n \to \mathbb{R}$ is a scalar-valued function. Assume that the functions Φ_k can *distinguish* the training points, in the sense that $\Phi_k(\mathbf{x}_\ell) = 0$ for $k \neq \ell$ and $\Phi_k(\mathbf{x}_k) = 1$. Show that the corresponding kernel matrix K defined in (11.99) is the identity matrix.

Solution: In this case,

$$\nabla_{\mathbf{w}} F(\mathbf{x};\mathbf{w}) = (\Phi_1(\mathbf{x}), \ldots, \Phi_m(\mathbf{x}))^T$$

and so the kernel matrix K has entries

$$k_{ij} = \sum_{k=1}^{m} \Phi_k(\mathbf{x}_i)\, \Phi_k(\mathbf{x}_j).$$

By the assumption that Φ_k *distinguishes* the training points, $k_{ij} = 1$ if $i = j$ and $k_{ij} = 0$ otherwise — that is, $K = \mathrm{I}$ is the identity matrix. ■